

## 明 細 書

システム、管理サーバ、引受サーバ、制御方法、制御プログラム、及び記録媒体

### 技術分野

[0001] 本発明は、システム、管理サーバ、引受サーバ、制御方法、制御プログラム、及び記録媒体に関する。特に、本発明は、管理サーバにより管理された複数の引受サーバが協調して、要求されたジョブを処理するグリッドコンピューティングシステム、管理サーバ、引受サーバ、制御方法、制御プログラム、及び記録媒体に関する。

### 背景技術

[0002] 近年、ネットワークで接続された複数のコンピュータが、ジョブに含まれる複数のステップ(処理部分)の何れかをそれぞれ処理することにより、ジョブを効率的に実行するグリッドコンピューティングが注目されている。グリッドコンピューティングによると、ジョブの依頼者は、ジョブの一部を処理するプログラムと、ジョブの処理に必要なデータとの組を、複数のコンピュータの各々に送信する。そして、複数のコンピュータの各々は、受け取ったデータを入力として、受け取ったプログラムを実行し、実行結果を依頼者に返信する。ジョブの依頼者は、返信された実行結果の各々を結合して、ジョブの結果を得る。これにより、ジョブに必要な計算処理量が膨大な場合であっても、ジョブの一部ずつを並列して複数のコンピュータにより実行させることにより、ジョブを効率的に完了させることができる。

非特許文献1には、グリッドコンピューティングシステムの一例が説明されている。  
非特許文献1:「情報処理」 社団法人情報処理学会発行、2003年発行、Vol.44、特集「PCグリッドの現在と展望」

### 発明の開示

#### 発明が解決しようとする課題

[0003] しかしながら、上記のグリッドコンピューティングによると、ジョブの依頼者のサーバのみが、複数のコンピュータの各々にプログラム及びデータを送信する。従って、依頼先のコンピュータの数に応じて、依頼者のサーバの負荷が増大すると共に、依頼

者のサーバへネットワークトラフィックが集中する。このため、多くのコンピュータにより効率的にジョブを完了させるには、多くの費用をかけて依頼者のサーバの設備を増強しなければならなかった。

- [0004] そこで本発明は、上記の課題を解決することのできるグリッドコンピューティングシステム、管理サーバ、引受サーバ、制御方法、制御プログラム、及び記録媒体を提供することを目的とする。この目的は特許請求の範囲における独立項に記載の特徴の組み合わせにより達成される。また従属項は本発明の更なる有利な具体例を規定する。

#### 課題を解決するための手段

- [0005] 上記課題を解決するために、利用者の要求するジョブに含まれる複数のステップを処理する複数の処理プログラムのうち、割り当てられた処理プログラムをそれぞれ実行する複数の引受サーバと、複数の引受サーバを管理する管理サーバとを備えたシステムであって、管理サーバは、複数の処理プログラムの各々を識別する識別情報と、複数の処理プログラムの実行順序とを含む実行指示情報を生成する実行指示情報生成部と、入力データを入力として最初に実行すべき処理プログラムを実行する引受サーバに、入力データと実行指示情報とを対応付けて送信する入力データ送信部とを有し、複数の引受サーバの各々は、受信した入力データを入力として割り当てられた前記処理プログラムを実行することにより入力データを更新する処理プログラム実行部と、入力データに対応付けて受信した実行指示情報に基づいて、更新された入力データを入力として次に実行すべき処理プログラムを実行させる引受サーバを選択する引受サーバ選択部と、引受サーバ選択部により選択された引受サーバに、受信した実行指示情報と更新された入力データとを対応付けて送信する入力データ受渡部とを有するシステム、当該システムの引受サーバ及び管理サーバ、当該引受サーバ又は管理サーバを制御する制御方法、当該引受サーバ又は管理サーバを制御する制御プログラム、及び制御プログラムを記録した記録媒体を提供する。

なお、上記の発明の概要は、本発明の必要な特徴の全てを列挙したものではなく、これらの特徴群のサブコンビネーションもまた、発明となりうる。

## 発明の効果

- [0006] 本発明によれば、より多くのコンピュータを結合してグリッドコンピューティングシステムを構築することができる。

## 発明を実施するための最良の形態

- [0007] 以下、発明の実施の形態を通じて本発明を説明するが、以下の実施形態は特許請求の範囲にかかる発明を限定するものではなく、また実施形態の中で説明されている特徴の組み合わせの全てが発明の解決手段に必須であるとは限らない。

- [0008] 図1は、グリッドコンピューティングシステム10の機能ブロック図である。グリッドコンピューティングシステム10は、LAN、WAN、又はインターネット等のネットワークにより互いに接続された、管理サーバ20と、利用者端末40と、引受サーバ50-1〜Nとを備える。利用者端末40は、例えば、パーソナル・コンピュータ等の端末であり、既存のグリッドコンピューティングシステムにおけるジョブの依頼者の端末と同様の構成により実現されてもよい。そして、引受サーバ50-1〜Nの各々には、ジョブに含まれる複数の処理部分を処理する複数の処理プログラムのうち、割り当てられた処理プログラムが予めインストールされて起動されている。データベース55-1〜Nの各々には、例えば同一のデータが予め格納されている。

- [0009] まず、利用者端末40は、利用者からの指示を受けて、管理サーバ20にジョブを要求する。管理サーバ20は、ジョブに入力される入力データと、複数の処理プログラムの実行順序等を示す実行指示情報との組を、送信用データ30として引受サーバ50-1に送る。引受サーバ50-1は、受け取った実行指示情報に基づいて、管理サーバ20から受け取った入力データを入力として処理プログラムを実行し、入力データを更新する。引受サーバ50-1は、更新した入力データと、実行指示情報との組とを、送信用データ30として引受サーバ50-2に送信する。

- [0010] この際、引受サーバ50-1は、入力データの次の送信先を、制御用データにより管理サーバ20に問合せてもよい。また、処理の内容によっては、引受サーバ50-1は、データベース55-1に格納されたデータにアクセスしてもよい。以降同様に、引受サーバ50-2〜(N-1)の各々は、入力データを入力として処理プログラムを順次実行する。最後に、引受サーバ50-Nは、全ての処理プログラムの実行が完了した完了

通知を管理サーバ20に送る。管理サーバ20は、受け取った完了通知を利用者端末40に転送する。

[0011] このように、本実施例に係るグリッドコンピューティングシステム10によると、各引受サーバは、処理プログラムを送信することなく、処理に必要なデータのみを送信する。この結果、多数のジョブが多数の利用者端末40等から入力された場合であっても、各ジョブの一部ずつをパイプライン方式で実行することにより、ジョブを効率的に実行することができる。

[0012] 図2は、管理サーバ20の機能ブロック図である。管理サーバ20は、実行指示情報生成部200と、入力データ送信部210と、引受サーバ情報格納部220と、問合せ対応部230と、処理プログラム起動部240と、障害発生通知送信部250と、引受サーバ変更部260と、削除指示部270とを有する。実行指示情報生成部200は、ジョブの要求を利用者端末40から受けると、ジョブに含まれる複数の処理部分を処理する複数の処理プログラムの各々を識別する識別情報と、その複数の処理プログラムの実行順序とを含む実行指示情報を生成する。そして、実行指示情報生成部200は、生成した実行指示情報を、ジョブに入力すべき入力データと共に入力データ送信部210に送る。

[0013] 入力データ送信部210は、入力データを入力として最初に実行すべき処理プログラムを実行する引受サーバ、例えば引受サーバ50-1に、入力データと実行指示情報とを対応付けて送信する。この際、入力データ送信部210は、最初に実行すべき処理プログラムを実行する引受サーバを、引受サーバ情報格納部220に格納されている情報に基づいて選択してもよい。

[0014] 引受サーバ情報格納部220は、複数の処理プログラムの各々を識別する識別情報と、その処理プログラムを実行する引受サーバの識別情報とを対応付けて格納する。更に、引受サーバ情報格納部220は、各引受サーバの識別情報に対応付けて、その引受サーバが新たに入力データを受信して処理プログラムを実行可能な実行可能状態か否かを示す実行可否情報を格納する。また、引受サーバ情報格納部220は、各引受サーバの識別情報に対応付けて、その引受サーバが入力データを受信した時刻を対応付けて更に格納してもよいし、その引受サーバの最大処理量及び計

算資源の使用率を対応付けて更に格納してもよい。

[0015] 問合せ対応部230は、引受サーバ50-1〜Nの何れかから、所定の処理プログラムを実行する引受サーバの問合せを受けた場合に、問合せを受けた処理プログラムの識別情報に対応する引受サーバの識別情報を引受サーバ情報格納部220から取得して、その問合せに対する返信として送信する。処理プログラム起動部240は、所定の条件が満たされる場合に、実行指示情報生成部200により生成された実行指示情報に基づいて、入力データを入力として2回目以降に実行すべき処理プログラムの各々を検出し、検出した処理プログラムの各々を、入力データ送信部210の送信先とは異なる引受サーバの何れかにおいて起動する。

[0016] 具体的には、処理プログラム起動部240は、2回目以降に実行すべき当該処理プログラムを実行する引受サーバが所定の基準台数より少ないという条件、又は、そのような引受サーバが1台も検出できないという条件が満たされる場合に、処理プログラムを起動してもよい。また、処理プログラム起動部240は、問合せ対応部230が問い合わせを受けた処理プログラムについて、その処理プログラムを実行する複数の引受サーバのうち、予め定められた基準割合を超える数の引受サーバが実行可能状態でない場合に、その処理プログラムを起動していない引受サーバの何れかにおいてその処理プログラム起動してもよい。

これにより、次の処理プログラムを予め準備できるので、ジョブの処理を中断されにくくすることができる。なお、これらの処理に伴い、処理プログラム起動部240は、引受サーバ情報格納部220のデータを更新してもよい。

[0017] 障害発生通知送信部250は、何れかの引受サーバから受信通知を受信してから予め定められた基準処理時間内に、その引受サーバから終了通知を受信したか否かを、引受サーバ情報格納部220に格納された情報に基づいて判断する。そして、障害発生通知送信部250は、基準処理時間内に受信できなかった場合に、当該引受サーバに入力データ等を送信した送信元の引受サーバに、当該引受サーバにおいて処理プログラムの実行に障害が発生した旨を示す障害発生通知を送信する。

[0018] 引受サーバ変更部260は、複数の引受サーバの何れかが処理プログラムを実行することにより使用する計算資源の使用率が、予め定められた基準使用率より低いかな

かを、引受サーバ情報格納部220に格納されたデータに基づいて判断する。そして、引受サーバ変更部260は、何れかの引受サーバにおいて計算資源の使用率が基準使用率より低い場合に、最大処理量はその引受サーバより小さい他の引受サーバに実行させるように、引受サーバ50-1〜Nを設定する。これに伴い、引受サーバ変更部260は、引受サーバ情報格納部220内のデータを更新する。

[0019] 削除指示部270は、最後に処理プログラムを実行した引受サーバ50-Nから完了通知を受けた場合に、複数の処理プログラムの全てが実行されることによりジョブが完了したと判断する。そしてこの場合、削除指示部270は、その複数の引き受けサーバの各々から、そのジョブの実行の過程で格納された入力データ及び実行指示情報を削除させる。これに代えて、削除指示部270は、これらの入力データ及び実行指示情報が不要である旨を通知するのみであってもよい。この場合、この通知を受けた引受サーバは、不要な入力データ及び実行指示情報を定期的に検出して削除してもよい。

[0020] 図3は、送信用データ30のデータ構造の一例を示す。(a)は、送信用データ30のデータ構造の概要を示す。送信用データ30は、本体部分と、ヘッダ部分とを含む。そして、送信用データ30は、本体部分に入力データを含み、ヘッダ部分に実行指示情報300を含む。このように、入力データ送信部210は、実行指示情報300を入力データのヘッダ部分に格納することにより、入力データ及び実行指示情報300を対応付けて送信してもよい。

[0021] 更に、好ましくは、送信用データ30は、実行指示情報300が示す処理プログラムのうち、既に実行を終えた処理プログラムを示す処理済フラグを、ヘッダ部分に含む。これにより、送信用データ30を受け取った引受サーバ50-1等は、複数の処理プログラムを実行可能な場合であっても、次にどの処理プログラムを実行すべきかを適切に判断できる。更に、送信用データ30は、処理済フラグに加えて、各処理プログラムの終了状態、例えば正常終了若しくは異常終了した旨、又はエラーコード等を示す情報を格納してもよい。これにより、障害発生時に障害の原因を調査しやすくすることができる。

[0022] これに代えて、送信用データ30は、既に実行を終えた処理プログラムの識別情報

を有しておらず、未だ実行されていない処理プログラムの識別情報のみを有していてもよい。即ち、引受サーバ50-1等は、既に実行を終えた処理プログラムの識別情報を送信用データ30から削除してもよい。これにより、送信用データ30のデータサイズを減少させることができる。

[0023] (b)は、実行指示情報300の一例を示す。実行指示情報300は、ジョブを識別する情報であるJOBを1行目に含む。そして、実行指示情報300は、第1のステップとして、識別情報がプログラムAである処理プログラムを実行すべき旨の指示を2行目に含

む。更に、実行指示情報300は、プログラムAの実行に用いるファイルを示す情報を

3行目及び4行目に含む。これらのファイルは、例えば、データベース55-1〜N等に格納されているファイルである。

[0024] 実行指示情報300は、第2のステップとして、識別情報がプログラムAである処理プログラムを実行すべき旨の指示を2行目に含む。第1のステップと同様に、実行指示情報300は、プログラムBの実行に用いるファイルを示す情報を、6行目及び7行目に含

む。なお、実行指示情報300は、例えば、ワークフロースクリプト又はJCL(Job Control Language)等の書式により記述されていてもよい。

[0025] ここで、処理プログラムとは、例えば、コンピュータのオペレーティングシステムで1つの実行可能ファイルとして取扱われるプログラムである。これに代えて、又はこれに加えて、処理プログラムとは、実行可能ファイル及び実行時に動的に読み込まれるライブラリファイルの組み合わせであってもよい。更に、処理プログラムとは、所定のプログラムが実行されることにより起動されるプロセス又はスレッドであってもよいし、所定のプログラム中に記述された少なくとも1つの関数、手続、メソッド、又はステップであってもよい。

[0026] このように、実行指示情報300は、処理プログラムの各々の識別情報、例えばプログラムA及びプログラムBと、これらの実行順序とを含んでいる。これにより、送信用データ30を受け取った引受サーバは、実行すべき処理プログラムを適切に選択でき

るのみならず、次に実行すべき処理プログラムを適切に選択することができる。

[0027] なお、実行指示情報300は、本図に示したデータの他に、ジョブの完了に必要な種々の情報を含んでいてもよい。例えば、実行指示情報300は、各処理プログラムが参照すべきデータベースの識別情報を更に含んでいてもよいし、全ての処理プログラムが実行された場合にその旨を通知すべき通知先を示す情報を更に含んでいてもよい。

[0028] 図4は、引受サーバ情報格納部220のデータ構造の一例を示す。引受サーバ情報格納部220は、各引受サーバの識別情報である引受サーバIDと、その引受サーバにインストールされて実行される処理プログラムの識別情報(ID)とを対応付けて格納する。更に、引受サーバ情報格納部220は、引受サーバIDに対応付けて、その引受サーバで処理中の入力データの数を示す処理数と、その引受サーバが他の引受サーバから入力データを受信した受信通知を最後に受信した時刻とを格納する。

[0029] ここで、引受サーバにおいて処理中の入力データの数とは、その引受サーバで実際に実行されている処理プログラムが入力とする入力データのみならず、その引受サーバに受信されたものの処理待ちで未処理である入力データの数をも含む。また、受信通知を最後に受信した時刻とは、受信通知を受信したものの、その受信通知に対応する終了通知を受け取っていない場合における受信通知の受信時刻をいう。これにより、障害発生通知送信部250は、各引受サーバから受信通知を受信してから終了通知を受信しないまま経過した時間を計測することができる。

[0030] 更に、引受サーバ情報格納部220は、各引受サーバIDに対応付けて、その引受サーバが新たに入力データを受信して処理プログラムを実行可能な実行可能状態か否かを示す実行可否情報を格納する。ここで、例えば、引受サーバ情報格納部220は、各引受サーバが許容する処理待ちの入力データの数と、その引受サーバで処理中の入力データとに基づいて、各引受サーバが実行可能状態か否かを判断して格納してもよい。例えば、ある引受サーバ(例えば、NodeBB)が許容する処理待ちの入力データの数が1つであり、その引受サーバで処理中の入力データの数が2つである場合には、引受サーバ情報格納部220は、その引受サーバを実行可能状態ではないと判断し、その旨の実行可否情報を格納する。



- [0031] また、引受サーバ情報格納部220は、各引受サーバIDに対応付けて、その引受サーバの最大処理量及び計算資源の使用率に対応付けて更に格納してもよい。ある引受サーバの最大処理量とは、例えば、その引受サーバのCPUの動作周波数であってもよいし、そのCPUの種類であってもよいし、そのCPUが単位時間あたりに処理可能なインストラクションの数であってもよい。これに加えて、最大処理量とは、その引受サーバのハードディスクドライブのアクセス速度若しくは容量、又は、その引受サーバのメモリのアクセス速度若しくは容量等により定まる値であってもよい。
- [0032] また、ある引受サーバにおける計算資源の使用率とは、例えば、所定の基準時間に対する、その基準時間内にその引受サーバのCPUが処理プログラムを実行している時間の割合である。これに代えて、使用率とは、メモリ又はハードディスクの占有率であってもよい。
- [0033] 図5は、引受サーバ50-2の機能ブロック図である。引受サーバ50-2は、処理プログラム実行部500と、引受サーバ選択部510と、キャッシュ部520と、入力データ受渡部530と、履歴格納部540と、障害発生判断部550と、変更復元部560とを有する。本図においては、引受サーバが、他の引受サーバから受信した入力データを更新し、更に他の引受サーバに受け渡す処理を説明するべく、引受サーバ50-2について説明する。なお、引受サーバ50-1及び引受サーバ50-3〜Nの各々は、本図に示す引受サーバ50-2と略同一の構成を取るので、説明を省略する。
- [0034] 処理プログラム実行部500は、実行可能状態である場合に引受サーバ50-1から入力データを受信すると、割り当てられた処理プログラムを例えばデータベース55-2から取得して、受信した入力データを入力として実行する。そして、処理プログラム実行部500は、処理プログラムの実行により入力データを更新し、更新した入力データを入力データ受渡部530に送る。また、処理プログラム実行部500は、処理プログラムの実行により、データベース55-2のデータを参照してもよいし、データベース55-2のデータを更新してもよい。
- [0035] なお、データベース55-1〜Nの各々は、互いに異なるデータを格納してもよいし、同一のデータを格納するクローンデータベースであってもよい。そしてこの場合、処理プログラム実行部500は、データベースの一貫性を保持する既存のシステムにより

、データベース55-2に対するデータの更新を、他のデータベース、例えばデータベース55-1及びデータベース55-3〜Nに反映してもよい。これにより、データベースの全てのデータを入力データとして送信する必要がなく、更新した部分のみが送信されるので、効率がよい。

[0036] 一方、処理プログラム実行部500は、実行可能状態でない場合に引受サーバ50-1から入力データを受信すると、新たに入力データを受け取れないと判断する。そして、処理プログラム実行部500は、入力データを拒絶する旨の拒絶通知を、入力データ及び実行指示情報の送信元の引受サーバ50-1に送信する。また、処理プログラム実行部500は、最後に入力データを受信してから予め定められた基準待機時間内に、入力データ及び実行指示情報を受信しなかった場合に、引受サーバ50-2における処理プログラムの実行を中止し、その旨を管理サーバ20に通知する。更に、処理プログラム実行部500は、処理プログラムの実行を中止する旨を管理サーバ20から指示された場合にも、処理プログラムの実行を中止する。

[0037] 引受サーバ選択部510は、引受サーバ50-1から実行指示情報を入力データに対応付けて受信する。そして、引受サーバ選択部510は、受信したその実行指示情報に基づいて、処理プログラム実行部500により更新された入力データを入力として次に実行すべき処理プログラムを実行させる引受サーバを選択する。例えば、引受サーバ選択部510は、更新された入力データを入力として次に実行すべき処理プログラムを実行する引受サーバの識別情報を、管理サーバ20に問合せ、問合せ結果の識別情報により識別される引受サーバを選択する。より具体的には、キャッシュ部520は、引受サーバ選択部510が過去に管理サーバ20の問合せ対応部230に問合せた結果取得した引受サーバの識別情報を、当該引受サーバが実行する処理プログラムの識別情報に対応付けてキャッシュしている。

[0038] そして、引受サーバ選択部510は、その処理プログラムを実行する引受サーバの識別情報がキャッシュ部520によりキャッシュされていない場合に、その処理プログラムを実行する引受サーバを管理サーバ20の問合せ対応部230に問合せる。一方、引受サーバ選択部510は、その処理プログラムを実行する引受サーバの識別情報がキャッシュ部520によりキャッシュされている場合に、識別情報のキャッシュされたその

引受サーバを選択する。また、引受サーバ選択部510は、障害発生判断部550により障害が発生したと判断された場合に、更新された入力データを入力として次に実行すべき処理プログラムを実行する他の引受サーバを再選択する。本図では、引受サーバ選択部510が、引受サーバ50-(N-1)を選択した場合について説明する。

[0039] 続いて、入力データ受渡部530は、引受サーバ選択部510により選択された引受サーバ50-(N-1)に、受信した実行指示情報と更新された入力データとを対応付けて送信する。そして、入力データ受渡部530は、送信した入力データ及び実行指示情報を履歴格納部540に格納する。そして、入力データ受渡部530は、引受サーバ50-(N-1)から拒絶通知を受けた場合に、その旨を障害発生判断部550に送る。この場合、入力データ受渡部530は、入力データ及び実行指示情報を履歴格納部540から取得して、引受サーバ選択部510により選択された他の引受サーバに送信する。

[0040] 履歴格納部540は、上記のように、入力データ受渡部530により引受サーバに送信された入力データ及び実行指示情報の履歴を格納する。更に、履歴格納部540は、処理プログラム実行部500が処理プログラムを実行することにより変更した引受サーバ50-2上のデータの履歴情報を、その変更を復元可能な情報に対応付けて格納する。ここで、引受サーバ50-2上のデータとは、例えば、データベース55-2に格納されているデータをいう。更に、引受サーバ50-2上のデータとは、引受サーバ50-2のみにより管理されているデータのみならず、ネットワーク・ファイル・システム又は共有ファイルシステムにより、他のサーバ装置と共有されている記憶装置のデータを含む。

[0041] 障害発生判断部550は、障害発生通知を管理サーバ20から受けると、入力データ受渡部530が更新された入力データ及び実行指示情報を送信した送信先の引受サーバにおいて、処理プログラムの実行に障害が発生したと判断する。他の例として、障害発生判断部550は、引受サーバ50-(N-1)から拒絶通知を受けた旨を入力データ受渡部530から通知された場合に、引受サーバ50-(N-1)において、処理プログラムの実行に障害が発生したと判断する。そして、障害発生判断部550は、障害の発生したジョブが、障害を回避して再試行可能な種類である場合には、他の引受サ

サーバを選択する旨を引受サーバ選択部510に通知し、当該種類でない場合には、処理プログラム実行部500によるデータの変更を元に戻す旨の復元指示を変更復元部560に通知する。

- [0042] 変更復元部560は、障害発生判断部550からの通知に基づいて、処理プログラム実行部500が変更したデータ、例えばデータベース55-2内のデータを、履歴格納部540の履歴情報に基づいて復元する。

以上、本図に示すように、引受サーバ50-2は、受信した入力データを入力として、予めインストールされ起動されている処理プログラムを実行すると共に、次に実行すべき処理プログラムを実行する引受サーバを選択し、その引受サーバに入力データを送信する。これにより、管理サーバ20における、処理プログラム及び入力データを送受信する負荷を軽減することができる。

- [0043] 以降、図6から図9を用いて、本実施例のグリッドコンピューティングシステム10において、管理サーバ20が引受サーバ50-1〜Nの各々に処理プログラムを実行させる4つの例について説明する。なお、図6から図9において、入力データ及び実行指示情報又は完了通知を点線で示し、その他の制御用データを一点破線で示す。

- [0044] 図6は、正常動作時におけるグリッドコンピューティングシステム10の動作フローの概要を示す(第1例)。利用者端末40は、管理サーバ20に対してジョブを要求する(S600)。そして、管理サーバ20は、実行指示情報を入力データのヘッダ部分に格納して、引受サーバ50-1に送信する(S610)。なお、管理サーバ20は、入力データを入力として2回目以降に実行すべき処理プログラムの各々を検出し、検出した処理プログラムを、引受サーバ50-1とは異なる引受サーバ、例えば引受サーバ50-Nにおいて起動してもよい。

- [0045] 引受サーバ50-1は、入力データを受信すると、受信通知を管理サーバ20に送信する(S620)。そして、引受サーバ50-1は、受信した入力データを入力として処理プログラム(例えば、プログラムA)を実行することにより入力データを更新する。処理プログラムの実行が終了すると、引受サーバ50-1は、終了通知を管理サーバ20に送信する。

- [0046] 続いて、引受サーバ50-1は、次に実行すべき処理プログラムを実行する引受サ

サーバを、必要に応じて管理サーバ20に問合せることにより選択する。引受サーバ50-1は、選択した引受サーバ50-2に、入力データ及び実行指示情報を送信する。管理サーバ20は、引受サーバ50-1から問合せを受けた場合に、問合せを受けた処理プログラムの識別情報に対応する引受サーバの識別情報を、引受サーバ50-1に返信する(S630)。

[0047] 引受サーバ50-2は、入力データを受信すると、受信通知を管理サーバ20に送信する(S640)。そして、引受サーバ50-2は、受信した入力データを入力として処理プログラム(例えば、プログラムB)を実行することにより入力データを更新する。処理プログラムの実行が終了すると、引受サーバ50-2は、終了通知を管理サーバ20に送信する。

[0048] 続いて、引受サーバ50-2は、次に実行すべき処理プログラムを実行する引受サーバを、必要に応じて管理サーバ20に問合せることにより選択する。引受サーバ50-2は、選択した引受サーバに、入力データ及び実行指示情報を送信する。管理サーバ20は、引受サーバ50-2から問合せを受けた場合に、問合せを受けた処理プログラムの識別情報に対応する引受サーバの識別情報を、引受サーバ50-2に返信する(S650)。

[0049] 続いて、引受サーバ50-2により送信された入力データ及び実行指示情報が、複数の引受サーバを順次経由して、引受サーバ50-Nに到達した場合について説明する。

引受サーバ50-Nは、受信した入力データを入力として処理プログラム(例えば、プログラムX)を実行することにより入力データを更新する(S660)。引受サーバ50-Nは、ジョブを処理する全ての処理プログラムの実行が完了した旨を示す完了通知を管理サーバ20に返信する。

[0050] 管理サーバ20は、受信した完了通知を利用者端末40に転送する(S670)。そして、管理サーバ20は、複数の引受サーバにおける履歴格納部から、入力データ及び実行指示情報を削除させる削除指示を、引受サーバ50-1〜Nの各々に送信する。

なお、本図で示したデータの送受信の他に、引受サーバ50-1〜Nの各々は、処理プログラムを実行することにより計算資源を使用する使用率を示す情報を、管理サ

サーバ20に定期的に送信してもよい。

[0051] 図7は、障害発生時におけるグリッドコンピューティングシステム10の動作フローの概要を示す(第2例)。本図のS600、S610、S620、及びS630における処理は、図6の同一符号のステップにおける処理と略同一であるので説明を省略する。S640において、引受サーバ50-2は、実行可能状態でないか、又は、処理プログラムの実行に障害が発生した場合に、入力データを拒絶する旨の拒絶通知を引受サーバ50-1に送信する(S640)。

[0052] 引受サーバ50-1は、入力データを入力として次に実行すべき処理プログラムを実行させる他の引受サーバを再選択する(S700)。例えば、引受サーバ50-1は、引受サーバ50-Nを選択する。そして、引受サーバ50-1は、入力データ及び実行指示情報を履歴格納部から取得して、引受サーバ50-Nに再送する。これを受けて、引受サーバ50-Nは、受信した入力データを入力として処理プログラムを実行することにより入力データを更新する(S660)。そして、引受サーバ50-Nは、次に実行すべき処理プログラムを実行する引受サーバを、必要に応じて管理サーバ20に問合せることにより選択する。S710における処理は、S630における処理と略同一であるので説明を省略する。

[0053] 図8は、障害発生時におけるグリッドコンピューティングシステム10の動作フローの概要を示す(第3例)。本図のS600、S610、S620、S630、S660、及びS700における処理は、図7の同一符号のステップにおける処理と略同一であるので説明を省略する。本図では、図7と異なり、S640において処理プログラムの実行に障害が発生し、かつ引受サーバ50-2が、拒絶通知を引受サーバ50-1に送信できなかった場合について説明する。

[0054] 管理サーバ20は、引受サーバ50-2から最後に受信通知を受信してから基準処理時間内に、引受サーバ50-2から終了通知を受信できなかった場合に、引受サーバ50-2において処理プログラムの実行に障害が発生したと判断する(S800)。そして、管理サーバ20は、その旨を示す障害発生通知を、引受サーバ50-2に入力データを送信した送信元、即ち引受サーバ50-1に送信する。

[0055] 引受サーバ50-1は、入力データを入力として次に実行すべき処理プログラムを

実行させる他の引受サーバを再選択する(S700)。例えば、引受サーバ50-1は、引受サーバ50-Nを選択する。そして、引受サーバ50-1は、入力データ及び実行指示情報を履歴格納部から取得して、引受サーバ50-Nに再送する。なお、S710における処理は、S630における処理と略同一であるので説明を省略する。

[0056] 図9は、障害発生時におけるグリッドコンピューティングシステム10の動作フローの概要を示す(第4例)。本図のS600、S610、S620、S630、S640、及びS650における処理は、図6の同一符号のステップにおける処理と略同一であるので説明を省略する。本図は、S660における処理プログラムの実行に障害が発生した場合について説明する。

[0057] 管理サーバ20は、引受サーバ50-Nから最後に受信通知を受信してから基準処理時間内に、引受サーバ50-Nから終了通知を受信できなかった場合に、引受サーバ50-Nにおいて処理プログラムの実行に障害が発生したと判断する(S800)。そして、管理サーバ20は、実行中のジョブが、障害を回避して再試行可能な種類でない場合には、引受サーバ50-N以前にそのジョブの処理プログラムを実行した全ての引受サーバに対して、処理プログラムの実行による変更を復元する復元指示を送る。

[0058] 引受サーバ50-1の変更復元部560は、この復元指示を受けると、入力データ受渡部530が更新された入力データ等を送信した送信先の何れかの引受サーバ、例えば引受サーバ50-2-Nの何れかに障害が発生したと判断する。そして、変更復元部560は、処理プログラム実行部500が変更したデータを、履歴格納部540の履歴情報に基づいて復元する(S900)。

[0059] 同様に、引受サーバ50-2の変更復元部560は、この復元指示を受けると、処理プログラム実行部500が変更したデータを、履歴格納部540の履歴情報に基づいて復元する(S910)。これにより、変更すべきデータの一部のみが変更され、データベース55-1又はデータベース55-2等のデータの一貫性が失われる事態を防ぐことができる。

[0060] 以降、図10から図16を用いて、図6から図9において説明した管理サーバ20及び引受サーバ50-1-Nの処理の詳細を説明する。なお、図6から図9と同様に、入力データ及び実行指示情報又は完了通知を点線で示し、その他の制御用データを一

点破線で示す。

- [0061] 図10は、S610の動作の詳細を示す。管理サーバ20は、利用者端末40からジョブの要求を受ける毎に、以下の処理を行う。まず、実行指示情報生成部200は、実行指示情報を生成する(S1000)。そして、処理プログラム起動部240は、実行指示情報に基づいて、入力データを入力として2回目以降に実行すべき処理プログラムを検出し、何れかの引受サーバにおいてその処理プログラムが既に起動されているか否かを判断する(S1010)。
- [0062] 引受サーバが1台も起動されていない場合に(S1010:NO)、処理プログラム起動部240は、その処理プログラムを起動する旨の指示を、例えば引受サーバ50-Nに送る(S1020)。この場合、処理プログラム起動部240は、新たに処理プログラムを起動させる引受サーバの識別情報を、その処理プログラムの識別情報に対応付けて、引受サーバ情報格納部220に格納する。
- [0063] これに代えて、処理プログラム起動部240は、過去に実行したジョブの情報に更に基づいて、以下の処理により、処理プログラムを新たに起動するか否かを判断してもよい。具体的には、処理プログラム起動部240は、各引受サーバの処理プログラム実行部500に、処理プログラムを実行させる毎に、その処理プログラムの実行に要した処理量を通知させる。この通知に基づいて、処理プログラム起動部240は、過去に受けたジョブの要求の種類に、その種類のジョブに含まれる複数の処理部分を処理する複数の処理プログラムの実行に要した処理量を対応付けて管理する。
- [0064] そして、処理プログラム起動部240は、ジョブの要求を受ける毎に、過去にその種類のジョブが実行された時に各処理プログラムの実行に要した処理量が、既にその処理プログラムを起動している引受サーバが処理可能な最大処理量を超える場合に、他の引受サーバにおいて処理プログラムを新たに起動する。これにより、不足する恐れのある処理プログラムを予め起動し、ジョブを円滑に完了できる。更に他の例として、処理プログラム起動部240は、ジョブの処理内容に基づいて、各処理プログラムの実行に要する処理量が、既にその処理プログラムを起動している引受サーバが処理可能な最大処理量を超えるか否かを判断してもよい。具体的には、処理プログラム起動部240は、処理プログラム毎に予め定められている処理量に基づいて判断して



もよいし、処理プログラムの実行内容を解析する処理を行ってもよい。

- [0065] 続いて、実行指示情報生成部200は、生成した実行指示情報を、ジョブに入力すべき入力データのヘッダ部分に格納することにより送信用データを生成し、例えば引受サーバ50-1に送信する(S1030)。
- [0066] 図11は、S620の動作の詳細を示す。引受サーバ50-1は、処理プログラムを起動している場合に、定期的に以下の処理を行う。まず、処理プログラム実行部500は、送信用データを受信したか否か判断する(S1100)。受信していない場合に(S1100:NO)、処理プログラム実行部500は、最後に送信用データを受信してから基準待機時間が経過したか否か判断する(S1110)。
- [0067] 基準待機時間が経過した場合、即ち、最後に送信用データを受信してから基準待機時間内に次の送信用データを受信できなかった場合に(S1110:YES)、処理プログラム実行部500は、処理プログラムを停止し(S1120)、処理プログラムを停止した旨を管理サーバ20に通知し(S1130)、処理を終了する。一方、処理プログラム実行部500は、送信用データを受信した場合に(S1100:YES)、引受サーバ50-1が実行可能状態か否か判断する(S1140)。実行可能状態でない場合に(S1140:NO)、処理プログラム実行部500は、拒絶通知を管理サーバ20に返信する(S1150)。ここで、実行可能状態の場合とは、処理待ち状態の入力データが所定以上である場合のみならず、引受サーバ50-1が既に処理プログラムAの実行を中止した場合も含む。
- [0068] これに加えて、処理プログラム実行部500は、以下の処理により拒絶通知を管理サーバ20に返信してもよい。具体的には、処理プログラム実行部500は、受信した実行指示情報に基づいて、受信した入力データを入力として既に実行されているべき処理プログラムを検出する。そして、処理プログラム実行部500は、検出したその処理プログラムが、入力データの送信元の引受サーバにより実行可能でなかった場合に、拒絶通知を返信する。例えば、処理プログラム実行部500は、検出したその処理プログラムが、入力データの送信元の引受サーバにより現在実行可能でない場合に、その引受サーバが入力データの送信時にもその処理プログラムを実行可能でなかったものと判断してもよい。

- [0069] 更にこれに加えて、処理プログラム実行部500は、入力データと共に受け取った電子書名の検証に失敗した場合に、拒絶通知を返信してもよい。これにより、送信元の引受サーバが正当な引受サーバに成りすます不正行為を防止することができる。
- [0070] 当該引受サーバが実行可能状態であり、検出したその処理プログラムが、入力データの送信元の引受サーバにより実行可能であって、かつ電子署名が正しく検証された場合に(S1140:YES)、処理プログラム実行部500は、受信通知を管理サーバ20に送信する(S1160)。そして、処理プログラム実行部500は、入力データを入力として割り当てられた処理プログラムを実行することによりその入力データを更新する(S1170)。履歴格納部540は、処理プログラム実行部500が処理プログラムを実行することにより変更した引受サーバ50-1上のデータの履歴情報を、その変更を復元可能な情報に対応付けて格納する(S1180)。
- [0071] 図12は、図11に続く動作の詳細を示す。引受サーバ選択部510は、管理サーバ20から受信した実行指示情報に基づいて、入力データを入力として次に実行すべき処理プログラムを検出する(S1200)。引受サーバ選択部510は、そのような処理プログラムを検出できない場合に(S1200:NO)、完了通知を管理サーバ20に送信する(S1205)。
- [0072] 一方、処理プログラムを検出できた場合に(S1200:YES)、引受サーバ選択部510は、キャッシュ部520を参照することにより、その処理プログラムの識別情報を実行する引受サーバの識別情報がキャッシュされているか否か判断する(S1210)。キャッシュされている場合に(S1220:YES)、引受サーバ選択部510は、キャッシュされているその識別情報の引受サーバを選択する(S1230)。
- [0073] 一方、キャッシュされていない場合に(S1220:NO)、引受サーバ選択部510は、その処理プログラムを実行する引受サーバの識別情報を管理サーバ20に問合せ、問合せ結果の識別情報により識別される引受サーバを選択する(S1240)。この場合、キャッシュ部520は、問合せ結果の識別情報を、問合せ対象の処理プログラムの識別情報に対応付けてキャッシュする(S1245)。
- [0074] 続いて、入力データ受渡部530は、引受サーバ選択部510により選択された引受サーバに、受信した実行指示情報と更新された入力データとを対応付けて送信用デ

ータとして送信する(S1250)。なお、入力データ受渡部530は、この更新された入力データ及び実行指示情報を対象として電子署名を作成し、作成した電子署名を、この入力データ及び実行指示情報に対応付けて送信してもよい。

[0075] そして、処理プログラム実行部500は、終了通知を管理サーバ20に送信する(S1260)。なお、処理プログラム実行部500は、処理プログラムの実行に所定以上の時間を要する場合等には、処理プログラムの処理経過を示す情報を、終了通知に先だって送信してもよい。続いて、履歴格納部540は、入力データ受渡部530により引受サーバに送信された入力データ及び実行指示情報の履歴を格納する(S1270)。なお、引受サーバ50-1は、本図に示す動作の他、処理プログラム実行部500が処理プログラムを実行することにより使用している計算資源の使用率を示す情報を、定期的に管理サーバ20に送ってもよい。

[0076] ここで、S640における動作は、送信用データの送信元が管理サーバ20に代えて引受サーバ50-1であり、当該送信先が引受サーバ50-2に代えて引受サーバ50-N等である他は、S620における動作と略同一であるので説明を省略する。同様に、S660における動作は、送信用データの送信元が管理サーバ20に代えて引受サーバ50-2である他は、S620における動作と略同一であるので説明を省略する。

[0077] 図13は、S630の動作の詳細を示す。管理サーバ20は、引受サーバ50-1〜Nの何れかから何らかの通知又は問合せを受ける毎に、以下の処理を行う。まず、引受サーバ情報格納部220は、受信通知又は終了通知を例えば引受サーバ50-1から受信したか否か判断する(S1300)。受信した場合に(S1300: YES)、引受サーバ情報格納部220は、受信通知又は終了通知に基づいて、各引受サーバで処理中のジョブの数を更新して、更新結果に基づいて実行可否情報を更新する(S1310)。

[0078] 続いて、問合せ対応部230は、所定の処理プログラムを実行する引受サーバの識別情報の問合せを受けたか否か判断する(S1320)。問合せを受けた場合に(S1320: YES)、処理プログラム起動部240は、問合せ対応部230が問い合わせを受けた処理プログラムについて、その処理プログラムを実行する複数の引受サーバのうち、実行可能状態でない引受サーバが、予め定められた基準割合を超えるか否か判断する(S1330)。

- [0079] 基準割合を超える場合に(S1330:YES)、処理プログラム起動部240は、その処理プログラムを起動していない引受サーバの何れかにおいてその処理プログラムを新たに起動する(S1340)。この場合、処理プログラム起動部240は、新たに処理プログラムを起動した引受サーバの識別情報を、その処理プログラムの識別情報に対応付けて、引受サーバ情報格納部220に格納する。
- [0080] そして、問合せ対応部230は、問合せを受けた処理プログラムの識別情報に対応する引受サーバの識別情報を引受サーバ情報格納部220から取得して、その問合せに対する返信として送信する(S1350)。引受サーバ情報格納部220が、その処理プログラムの識別情報に対応する引受サーバの識別情報を複数格納している場合には、問合せ対応部230は、その処理プログラムの識別情報に対応する全ての識別情報を送信してもよいし、所定の判断基準に基づいて一部の識別情報を選択して送信してもよい。
- [0081] 例えば、問合せ対応部230は、問合せた引受サーバとの間の通信速度がより速い引受サーバの識別情報を、問合せた引受サーバとの間の通信速度がより遅い引受サーバの識別情報より優先して選択し、送信してもよい。また、S1340において処理プログラム起動部240により新たに処理プログラムが起動された場合には、問合せ対応部230は、処理プログラムが起動されたその引受サーバの識別情報のみを送信してもよい。
- [0082] 更に、引受サーバ情報格納部220が、引受サーバ50-1〜Nの何れかから、処理プログラムを実行することにより使用する計算資源の使用率を受信した場合に(S1360:YES)、引受サーバ変更部260は、その使用率が基準使用率以下か否かを判断する(S1370)。基準使用率以下の場合に(S1370:YES)、引受サーバ変更部260は、その処理プログラムを、その使用率を送信した引受サーバより最大処理量が少ない他の引受サーバに実行させる(S1380)。具体的には、引受サーバ変更部260は、その処理プログラムを実行していた引受サーバに、その処理プログラムの実行を中止する中止指示を送信し、その引受サーバより最大処理量が少ない当該他の引受サーバに、その処理プログラムを起動する起動指示を送信する。
- [0083] この結果、変更前の引受サーバは、処理プログラム実行部500により当該処理プロ

プログラムの実行を停止し、変更後の引受サーバは、処理プログラム実行部500により当該処理プログラムを起動する。そしてこの場合、処理プログラム起動部240は、新たに処理プログラムを起動した引受サーバの識別情報を、その処理プログラムの識別情報に対応付けて、引受サーバ情報格納部220に格納する。一方、処理プログラム起動部240は、処理プログラムの実行を中止した引受サーバの識別情報を、引受サーバ情報格納部220から削除する。

[0084] なお、S650における動作は、問合せ元の引受サーバが引受サーバ50-1に代えて引受サーバ50-2である事項の他は、S630における動作と略同一であるので説明を省略する。

[0085] 図14は、S670の動作の詳細を示す。削除指示部270は、最後に処理プログラムを実行した引受サーバ50-Nから完了通知を受けた場合に(S1400: YES)、複数の処理プログラムの全てが実行されることによりジョブが完了したと判断し、その旨を利用者端末40に通知する(S1410)。そして、削除指示部270は、引受サーバ50-1-Nの各々に指示して、引受サーバ50-1-Nの各々における履歴格納部540から、そのジョブの実行の過程で格納された入力データ、実行指示情報、及び履歴情報を削除する(S1420)。

[0086] 図15は、S700の動作の詳細を示す。障害発生判断部550は、入力データの送信先の引受サーバで障害が発生したか否か判断する(S1500)。例えば、障害発生判断部550は、障害発生通知を管理サーバ20から受けるか、又は拒絶通知を引受サーバ50-2から受けた場合に、障害が発生したと判断する。

[0087] そして、障害発生判断部550により障害が発生したと判断された場合に(S1500: YES)、引受サーバ選択部510は、実行中に障害が発生したその処理プログラムを実行する他の引受サーバの識別情報を管理サーバ20に問合せする(S1510)。この場合、キャッシュ部520は、問合せ結果の識別情報を、問合せ対象の処理プログラムの識別情報に対応付けてキャッシュする(S1520)。

[0088] 引受サーバ選択部510は、問合せ結果の識別情報により識別される引受サーバを選択する(S1530)。例えば、問合せ対応部230は、当該引受サーバとの通信速度等がより速い引受サーバの識別情報を優先して問合せ結果として返信する。この結果、

引受サーバ選択部510は、更新された入力データを入力として次に実行すべき処理プログラムを実行している複数の引受サーバのうち、当該引受サーバとの間の通信速度がより速い引受サーバを、当該通信速度がより遅い引受サーバより優先して選択する。一例として、引受サーバ選択部510は、当該通信速度が最も速い引受サーバを選択し、他の引受サーバを選択しない。

[0089] 入力データ受渡部530は、既に送信した入力データ及び実行指示情報を履歴格納部540から取得する(S1540)。入力データ受渡部530は、取得したこれらの入力データ及び実行指示情報を、送信用データとして、再選択した引受サーバ、例えば引受サーバ50-Nに送信する(S1550)。

[0090] 図16は、S800の動作の詳細を示す。管理サーバ20は、ジョブを引受サーバ50-1-Nに実行させている場合に、例えば定期的に以下の処理を繰り返す。まず、引受サーバ情報格納部220は、各引受サーバについて、その引受サーバから最後に受信通知を受信してから基準処理時間が経過したか否かを判断する(S1600)。基準処理時間が経過した場合に、即ち、基準処理時間内に終了通知を受信できなかった場合に(S1600:YES)、引受サーバ情報格納部220は、障害の発生したジョブが再試行可能な種類であるか否かを判断する(S1610)。

[0091] 再試行可能な種類である場合に(S1610:YES)、引受サーバ情報格納部220は、障害発生通知を引受サーバ50-1に送信する(S1620)。更に、引受サーバ情報格納部220は、入力データの送信先の引受サーバのハードウェアに障害が発生した場合等には、障害の発生したその引受サーバに対応付けて格納されていた処理プログラムの識別情報を削除してもよい。一方、再試行可能でない種類である場合に(S1610:NO)、引受サーバ情報格納部220は、そのジョブの一部を実行した処理プログラムによるデータの変更を復元する旨の指示を、すでに処理プログラムを実行した引受サーバに送る。例えば、引受サーバ情報格納部220は、引受サーバ50-1及び引受サーバ50-2にその指示を送る(S1630)。

[0092] 図17は、管理サーバ20として機能するコンピュータのハードウェア構成の一例を示す。管理サーバ20は、ホストコントローラ1782により相互に接続されるCPU1700、RAM1720、グラフィックコントローラ1775、及び表示装置1780を有するCPU周辺

部と、入出力コントローラ1784によりホストコントローラ1782に接続される通信インターフェイス1730、ハードディスクドライブ1740、及びCD-ROMドライブ1760を有する入出力部と、入出力コントローラ1784に接続されるBIOS1710、フレキシブルディスクドライブ1750、及び入出力チップ1770を有するレガシー入出力部とを備える。

- [0093] ホストコントローラ1782は、RAM1720と、高い転送レートでRAM1720をアクセスするCPU1700及びグラフィックコントローラ1775とを接続する。CPU1700は、BIOS1710及びRAM1720に格納されたプログラムに基づいて動作し、各部の制御を行う。グラフィックコントローラ1775は、CPU1700等がRAM1720内に設けたフレームバッファ上に生成する画像データを取得し、表示装置1780上に表示させる。これに代えて、グラフィックコントローラ1775は、CPU1700等が生成する画像データを格納するフレームバッファを、内部に含んでもよい。
- [0094] 入出力コントローラ1784は、ホストコントローラ1782と、比較的高速な入出力装置である通信インターフェイス1730、ハードディスクドライブ1740、及びCD-ROMドライブ1760を接続する。通信インターフェイス1730は、ネットワークを介して外部の装置と通信する。ハードディスクドライブ1740は、管理サーバ20が使用するプログラム及びデータを格納する。CD-ROMドライブ1760は、CD-ROM1795からプログラム又はデータを読み取り、RAM1720を介して入出力チップ1770に提供する。
- [0095] また、入出力コントローラ1784には、BIOS1710と、フレキシブルディスクドライブ1750や入出力チップ1770等の比較的低速な入出力装置とが接続される。BIOS1710は、管理サーバ20の起動時にCPU1700が実行するブートプログラムや、管理サーバ20のハードウェアに依存するプログラム等を格納する。フレキシブルディスクドライブ1750は、フレキシブルディスク1790からプログラム又はデータを読み取り、RAM1720を介して入出力チップ1770に提供する。入出力チップ1770は、フレキシブルディスク1790や、例えばパラレルポート、シリアルポート、キーボードポート、マウスポート等を介して各種の入出力装置を接続する。
- [0096] 管理サーバ20に提供されるプログラムは、フレキシブルディスク1790、CD-ROM1795、又はICカード等の記録媒体に格納されて利用者によって提供される。プログ

ラムは、入出力チップ1770及び／又は入出力コントローラ1784を介して、記録媒体から読み出され管理サーバ20にインストールされて実行される。

[0097] 管理サーバ20にインストールされて実行される制御プログラムは、実行指示情報生成モジュールと、入力データ送信モジュールと、引受サーバ情報格納モジュールと、問合せ対応モジュールと、処理プログラム起動モジュールと、障害発生通知送信モジュールと、引受サーバ変更モジュールと、削除指示モジュールとを含む。各モジュールが管理サーバ20に働きかけて行わせる動作は、図1から図16において説明した管理サーバ20における、対応する部材の動作と同一であるから、説明を省略する。

[0098] また、管理サーバ20に提供される制御プログラムは、ネットワークなどを介して引受サーバ50-1-Nの各々にインストールされて実行されてもよい。引受サーバ50-1-Nの各々にインストールされて実行される制御プログラムは、処理プログラム実行モジュールと、引受サーバ選択モジュールと、キャッシュモジュールと、入力データ受渡モジュールと、履歴格納モジュールと、障害発生判断モジュールとを含む。各モジュールが引受サーバ50-1-Nの各々に働きかけて行わせる動作は、図1から図16において説明した引受サーバ50-1-Nの各々における、対応する部材の動作と同一であるから、説明を省略する。

[0099] 以上を示したプログラム又はモジュールは、外部の記憶媒体に格納されてもよい。記憶媒体としては、フレキシブルディスク1790、CD-ROM1795の他に、DVDやPD等の光学記録媒体、MD等の光磁気記録媒体、テープ媒体、ICカード等の半導体メモリ等を用いることができる。また、専用通信ネットワークやインターネットに接続されたサーバシステムに設けたハードディスク又はRAM等の記憶装置を記録媒体として使用し、ネットワークを介してプログラムを管理サーバ20に提供してもよい。

[0100] 以上、本実施例に示すように、管理サーバ20は、処理プログラムを引受サーバに実行させる毎に入力データを送受信することなく、入力データを複数の引受サーバに順次受渡しさせることができる。この結果、管理サーバ20の負荷が軽減されるとともに、管理サーバ20へのネットワークトラフィックが減少する。更に、引受サーバ50は、入力データの送信先をキャッシュしており、キャッシュされていない送信先のみを管理サーバ20に問合せする。この結果、管理サーバ20へのネットワークトラフィックが更



に減少する。

- [0101] 更に、本実施例によると、処理プログラムの実行に障害が発生した場合であっても、引受サーバ50は、できるだけ管理サーバ20に問合せずに障害の復旧を試みる。この結果、管理サーバ20へのネットワークトラフィックを減少できる。また、管理サーバ20は、実行させる処理プログラムを、引受サーバに予めインストールして起動させておく。これにより、管理サーバ20のみが、入力データにアクセスすることを許可する引受サーバを管理していればよく、引受サーバにそのような管理をさせる負担を軽減できる。また、管理サーバ20は、引受サーバ50からの問合せ等に応じて、各引受サーバの負荷を分散することができる。

以上、本実施例によると、より多くのコンピュータを結合して適切かつ柔軟なグリッドコンピューティングシステムを構築することができる。

- [0102] 以上、本発明を実施の形態を用いて説明したが、本発明の技術的範囲は上記実施の形態に記載の範囲には限定されない。上記実施の形態に、多様な変更または改良を加えることが可能であることが当業者に明らかである。その様な変更または改良を加えた形態も本発明の技術的範囲に含まれ得ることが、特許請求の範囲の記載から明らかである。

#### 図面の簡単な説明

- [0103] [図1]図1は、グリッドコンピューティングシステム10の機能ブロック図である。
- [図2]図2は、管理サーバ20の機能ブロック図である。
- [図3]図3は、送信用データ30のデータ構造の一例を示す。
- [図4]図4は、引受サーバ情報格納部220のデータ構造の一例を示す。
- [図5]図5は、引受サーバ50-2の機能ブロック図である。
- [図6]図6は、正常動作時におけるグリッドコンピューティングシステム10の動作フローの概要を示す(第1例)。
- [図7]図7は、障害発生時におけるグリッドコンピューティングシステム10の動作フローの概要を示す(第2例)。
- [図8]図8は、障害発生時におけるグリッドコンピューティングシステム10の動作フローの概要を示す(第3例)。

[図9]図9は、障害発生時におけるグリッドコンピューティングシステム10の動作フローの概要を示す(第4例)。

[図10]図10は、S610の動作の詳細を示す。

[図11]図11は、S620の動作の詳細を示す。

[図12]図12は、図11に続く動作の詳細を示す。

[図13]図13は、S630の動作の詳細を示す。

[図14]図14は、S670の動作の詳細を示す。

[図15]図15は、S700の動作の詳細を示す。

[図16]図16は、S800の動作の詳細を示す。

[図17]図17は、管理サーバ20として機能するコンピュータのハードウェア構成の一例を示す。

#### 符号の説明

[0104] 10 グリッドコンピューティングシステム

20 管理サーバ

30 送信用データ

40 利用者端末

50 引受サーバ

55 データベース

200 実行指示情報生成部

210 入力データ送信部

220 引受サーバ情報格納部

230 問合対応部

240 処理プログラム起動部

250 障害発生通知送信部

260 引受サーバ変更部

270 削除指示部

300 実行指示情報

500 処理プログラム実行部

510 引受サーバ選択部

520 キャッシュ部

530 入力データ受渡部

540 履歴格納部

550 障害発生判断部

560 変更復元部

### 請求の範囲

- [1] 利用者の要求するジョブに含まれる複数のステップを処理する複数の処理プログラムのうち、割り当てられた処理プログラムをそれぞれ実行する複数の引受サーバと、前記複数の引受サーバを管理する管理サーバとを備えたシステムであって、  
前記管理サーバは、  
前記複数の処理プログラムの各々を識別する識別情報と、前記複数の処理プログラムの実行順序とを含む実行指示情報を生成する実行指示情報生成部と、  
前記ジョブに入力される入力データを入力として最初に実行すべき処理プログラムを実行する引受サーバに、前記入力データと前記実行指示情報とを対応付けて送信する入力データ送信部と  
を有し、  
前記複数の引受サーバの各々は、  
受信した前記入力データを入力として割り当てられた前記処理プログラムを実行することにより前記入力データを更新する処理プログラム実行部と、  
前記入力データに対応付けて受信した実行指示情報に基づいて、更新された前記入力データを入力として次に実行すべき処理プログラムを実行させる引受サーバを選択する引受サーバ選択部と、  
前記引受サーバ選択部により選択された前記引受サーバに、受信した前記実行指示情報と更新された前記入力データとを対応付けて送信する入力データ受渡部とを有するシステム。
- [2] 前記引受サーバ選択部は、更新された前記入力データを入力として次に実行すべき処理プログラムを実行する引受サーバの識別情報を、前記管理サーバに問合せ、問合せ結果の識別情報により識別される引受サーバを選択し、  
前記管理サーバは、  
複数の処理プログラムの各々を識別する識別情報と、当該処理プログラムを実行する引受サーバの識別情報とを対応付けて格納する引受サーバ情報格納部と、  
前記引受サーバ選択部から問合せを受けた場合に、問合せを受けた処理プログラムの識別情報に対応する引受サーバの識別情報を、前記引受サーバ情報格納部か

ら取得して、前記問合せに対する返信として送信する問合せ対応部とを更に有する請求項1記載のシステム。

- [3] 前記複数の引受サーバの各々は、前記引受サーバ選択部が過去に前記問合せ対応部に問合せた結果取得した引受サーバの識別情報を、当該引受サーバが実行する処理プログラムの識別情報に対応付けてキャッシュするキャッシュ部を更に備え、

前記引受サーバ選択部は、実行すべき処理プログラムの識別情報が、前記キャッシュ部によりキャッシュされている場合に、当該処理プログラムに対応する引受サーバを選択し、実行すべき処理プログラムの識別情報が、前記キャッシュ部によりキャッシュされていない場合に、当該処理プログラムを実行する引受サーバを前記問合せ対応部に問合せる 請求項2記載のシステム。

- [4] 前記処理プログラム実行部は、前記入力データを受信した場合に、入力データを受信した旨を示す受信通知を前記管理サーバに送信し、前記処理プログラムの実行が終了した場合に、当該処理プログラムの実行が終了した旨を示す終了通知を前記管理サーバに送信し、

前記引受サーバ情報格納部は、前記複数の引受サーバの各々が新たに入力データを受信して処理プログラムを実行可能か否かを、前記受信通知及び前記終了通知に基づいて判断した判断結果を実行可否情報として更に格納し、

前記問合せ対応部は、問合せを受けた処理プログラムについて、当該処理プログラムを実行する複数の引受サーバのうち、新たに入力データを受信して処理プログラムを実行可能な引受サーバを、前記実行可否情報に基づいて選択し、選択した当該引受サーバの識別情報を前記問合せに対する返信として送信する

請求項2記載のシステム。

- [5] 前記管理サーバは、前記問合せ対応部が問合せを受けた処理プログラムについて、当該処理プログラムを実行する複数の引受サーバのうち、予め定められた基準割合を超える数の引受サーバが、新たに入力データを受信して処理プログラムを実行可能でない場合に、当該処理プログラムを起動していない引受サーバの何れかにおいて当該処理プログラムを起動する処理プログラム起動部を更に備え、

前記問合せ対応部は、前記処理プログラム起動部により当該処理プログラムが起動さ

れた引受サーバの識別情報を、前記問合せに対する返信として送信する  
請求項4記載のシステム。

- [6] 前記複数の引受サーバの各々は、  
前記入力データ受渡部により引受サーバに送信された入力データ及び実行指示  
情報の履歴を格納する履歴格納部と、  
前記入力データ受渡部が更新された前記入力データ及び実行指示情報を送信し  
た送信先の引受サーバにおいて、処理プログラムの実行に障害が発生したか否かを判  
断する障害発生判断部と  
を更に有し、  
前記障害発生判断部により障害が発生したと判断された場合において、  
前記引受サーバ選択部は、更新された前記入力データを入力として次に実行する  
べき処理プログラムを実行する他の引受サーバを選択し、  
前記入力データ受渡部は、前記履歴格納部から入力データ及び実行指示情報を  
取得して、前記引受サーバ選択部により選択された前記他の引受サーバに送信する  
請求項1記載のシステム。
- [7] 前記処理プログラム実行部は、前記入力データを受信した場合に、入力データを  
受信した旨を示す受信通知を前記管理サーバに送信し、前記処理プログラムの実行  
が終了した場合に、当該処理プログラムの実行が終了した旨を示す終了通知を前記  
管理サーバに送信し、  
前記管理サーバは、何れかの引受サーバから前記受信通知を受信してから予め定  
められた基準処理時間内に、当該引受サーバから前記終了通知を受信できなかった  
場合に、当該引受サーバに入力データを送信した送信元の引受サーバに、当該  
引受サーバにおいて処理プログラムの実行に障害が発生した旨を示す障害発生通  
知を送信する障害発生通知送信部を更に備え、  
前記障害発生判断部は、前記障害発生通知を受信した場合に、前記入力データ  
受渡部が更新された前記入力データ及び実行指示情報を送信した送信先の引受サ  
ーバにおいて処理プログラムの実行に障害が発生したと判断する  
請求項6記載のシステム。

- [8]      入力データ及び実行指示情報を送信した送信先の引受サーバにおいて、  
前記処理プログラム実行部は、新たに入力データを受信して処理プログラムを実行  
可能でないと判断した場合に、入力データを拒絶する旨を示す拒絶通知を当該入力  
データ及び当該実行指示情報の送信元の引受サーバに送信し、  
当該送信元の引受サーバにおける前記障害発生判断部は、当該送信先の引受サ  
ーバから前記拒絶通知を受信した場合に、当該送信先の引受サーバにおいて処理  
プログラムの実行に障害が発生したと判断する  
請求項6記載のシステム。
- [9]      前記管理サーバは、前記複数の処理プログラムが実行されることにより前記ジョブ  
が完了した場合に、前記複数の引受サーバの各々における前記履歴格納部から、  
入力データ及び実行指示情報を削除させる削除指示部を更に備える  
請求項6記載のシステム。
- [10]     前記複数の引受サーバの各々において、  
前記引受サーバ選択部は、更新された前記入力データを入力として次に実行する  
べき処理プログラムを起動している複数の引受サーバのうち、当該引受サーバとの間  
の通信速度がより速い引受サーバを、当該通信速度がより遅い引受サーバより優先  
して、前記他の引受サーバとして選択する  
請求項6記載のシステム。
- [11]     前記複数の引受サーバの各々は、  
前記処理プログラム実行部が前記処理プログラムを実行することにより変更した当  
該引受サーバ上のデータの履歴情報を、当該変更を復元可能な情報に対応付けて  
格納する履歴格納部と、  
前記入力データ受渡部が更新された前記入力データ及び実行指示情報を送信し  
た送信先の引受サーバにおいて、処理プログラムの実行に障害が発生した場合に、  
前記処理プログラム実行部が変更したデータを、前記履歴格納部の前記履歴情報  
に基づいて復元する変更復元部と  
を更に有する  
請求項1記載のシステム。

- [12] 前記管理サーバは、生成した前記実行指示情報に基づいて、前記ジョブに入力される前記入力データを入力として2回目以降に実行すべき処理プログラムの各々を検出し、検出した処理プログラムの各々を、前記入力データ送信部の送信先とは異なる引受サーバの何れかにおいて起動する処理プログラム起動部を更に備える請求項1記載のシステム。
- [13] 前記複数の引受サーバの各々において、  
前記処理プログラム実行部は、過去に処理プログラムの実行に要した処理量を前記管理サーバに通知し、  
前記管理サーバにおいて、  
前記処理プログラム起動部は、前記処理プログラム実行部から通知された処理プログラムの過去の処理量が、既に当該処理プログラムを起動している引受サーバが処理可能な最大処理量を超える場合に、前記入力データ送信部の送信先とは異なる引受サーバの何れかにおいて、当該処理プログラムを起動する  
請求項12記載のシステム。
- [14] 前記管理サーバは、前記ジョブに入力される前記入力データを入力として実行すべき複数の処理プログラムの各々を検出し、検出した処理プログラムの各々について、当該入力データを入力として実行する当該処理プログラムに必要な処理量が、既に当該処理プログラムを起動している引受サーバが処理可能な最大処理量を超える場合に、当該処理プログラムを、前記入力データ送信部の送信先とは異なる引受サーバの何れかにおいて起動する処理プログラム起動部  
を更に備える請求項1記載のシステム。
- [15] 前記処理プログラム実行部は、最後に入力データ及び実行指示情報を受信してから予め定められた基準待機時間内に、入力データ及び実行指示情報を受信しなかった場合に、当該引受サーバの処理プログラムの実行を中止する  
請求項1記載のシステム。
- [16] 前記管理サーバは、前記複数の引受サーバの何れかにおいて、前記処理プログラム実行部が処理プログラムを実行することにより使用する計算資源の使用率が、予め定められた基準使用率より低い場合に、当該処理プログラムを、最大処理量が当該



引受サーバより少ない他の引受サーバに実行させる引受サーバ変更部を更に備える

請求項1記載のシステム。

- [17] 入力データ及び実行指示情報を送信した送信先の引受サーバにおいて、  
前記処理プログラム実行部は、受信した実行指示情報に基づいて、受信した入力データに既に実行されているべき処理プログラムを検出し、検出した当該処理プログラムが送信元の引受サーバにより実行可能であった場合に、処理プログラムを実行する

請求項1記載のシステム。

- [18] 前記複数の引受サーバの各々において、  
前記入力データ受渡部は、更新された前記入力データ又は前記実行指示情報の電子署名を作成し、作成した当該電子署名を、電子署名の対象である当該入力データ及び当該実行指示情報に更に対応付けて送信し、  
入力データ、実行指示情報、及び電子署名を受信した引受サーバにおいて、  
前記処理プログラム実行部は、前記電子署名が正しく検証された場合に処理プログラムを実行する

請求項1記載のシステム。

- [19] 利用者の要求するジョブに含まれる複数のステップを処理する複数の処理プログラムのうち、割り当てられた処理プログラムをそれぞれ実行する複数の引受サーバを管理する管理サーバであって、  
前記複数の引受サーバの各々は、  
受信した入力データを入力として、割り当てられた前記処理プログラムを実行することにより、前記入力データを更新する処理プログラム実行部と、  
前記入力データに対応付けて受信した実行指示情報に基づいて、更新された前記入力データを入力として次に実行するべき処理プログラムを実行させる他の引受サーバを選択する引受サーバ選択部と、  
前記引受サーバ選択部により選択された前記引受サーバに、受信した前記実行指示情報と更新された前記入力データとを対応付けて送信する入力データ受渡部と

を有し、

前記複数の処理プログラムの各々を識別する識別情報と、前記複数の処理プログラムの実行順序とを含む実行指示情報を生成する実行指示情報生成部と、

前記ジョブに入力される入力データを入力として最初に実行すべき処理プログラムを実行する引受サーバに、当該入力データと前記実行指示情報とを対応付けて送信する入力データ送信部と  
を備える管理サーバ。

- [20] 管理サーバにより管理され、利用者の要求するジョブに含まれる複数のステップを処理する複数の処理プログラムのうち、割り当てられた処理プログラムを実行する引受サーバであって、

前記管理サーバは、

前記複数の処理プログラムの各々を識別する識別情報と、前記複数の処理プログラムの実行順序とを含む実行指示情報を生成する実行指示情報生成部と、

前記ジョブに入力される入力データを入力として最初に実行すべき処理プログラムを実行する引受サーバに、前記入力データと前記実行指示情報とを対応付けて送信する入力データ送信部と

を有し、

受信した前記入力データを入力として、割り当てられた前記処理プログラムを実行することにより前記入力データを更新する処理プログラム実行部と、

前記入力データに対応付けて受信した実行指示情報に基づいて、更新された前記入力データを入力として次に実行すべき処理プログラムを実行させる他の引受サーバを選択する引受サーバ選択部と、

前記引受サーバ選択部により選択された前記引受サーバに、受信した前記実行指示情報と更新された前記入力データとを対応付けて送信する入力データ受渡部と  
を備える引受サーバ。

- [21] 利用者の要求するジョブに含まれる複数のステップを処理する複数の処理プログラムのうち、割り当てられた処理プログラムをそれぞれ実行する複数の引受サーバを管理する管理サーバを制御する制御方法であって、前記複数の引受サーバの各々

は、

受信した入力データを入力として、割り当てられた前記処理プログラムを実行することにより、前記入力データを更新する処理プログラム実行部と、

前記入力データに対応付けて受信した実行指示情報に基づいて、更新された前記入力データを入力として次に実行すべき処理プログラムを実行させる他の引受サーバを選択する引受サーバ選択部と、

前記引受サーバ選択部により選択された前記引受サーバに、受信した前記実行指示情報と更新された前記入力データとを対応付けて送信する入力データ受渡部とを有し、

前記複数の処理プログラムの各々を識別する識別情報と、前記複数の処理プログラムの実行順序とを含む実行指示情報を生成する実行指示情報生成段階と、

前記ジョブに入力される入力データを入力として最初に実行すべき処理プログラムを実行する引受サーバに、当該入力データと前記実行指示情報とを対応付けて送信する入力データ送信段階と

を備える制御方法。

- [22] 管理サーバにより管理され、利用者の要求するジョブに含まれる複数のステップを処理する複数の処理プログラムのうち、割り当てられた処理プログラムを実行する引受サーバを制御する制御方法であって、

前記管理サーバは、

前記複数の処理プログラムの各々を識別する識別情報と、前記複数の処理プログラムの実行順序とを含む実行指示情報を生成する実行指示情報生成部と、

前記ジョブに入力される入力データを入力として最初に実行すべき処理プログラムを実行する引受サーバに、前記入力データと前記実行指示情報とを対応付けて送信する入力データ送信部と

を有し、

受信した前記入力データを入力として、割り当てられた前記処理プログラムを実行することにより前記入力データを更新する処理プログラム実行段階と、

前記入力データに対応付けて受信した実行指示情報に基づいて、更新された前

記入入力データを入力として次に実行すべき処理プログラムを実行させる他の引受サーバを選択する引受サーバ選択段階と、

前記引受サーバ選択段階において選択された前記引受サーバに、受信した前記実行指示情報と更新された前記入力データとを対応付けて送信する入力データ受渡段階と

を備える制御方法。

- [23] 利用者の要求するジョブに含まれる複数のステップを処理する複数の処理プログラムのうち、割り当てられた処理プログラムをそれぞれ実行する複数の引受サーバを管理する管理サーバとして、コンピュータを機能させる制御プログラムであって、

前記複数の引受サーバの各々は、

受信した入力データを入力として、割り当てられた前記処理プログラムを実行することにより、前記入力データを更新する処理プログラム実行部と、

前記入力データに対応付けて受信した実行指示情報に基づいて、更新された前記入力データを入力として次に実行すべき処理プログラムを実行させる他の引受サーバを選択する引受サーバ選択部と、

前記引受サーバ選択部により選択された前記引受サーバに、受信した前記実行指示情報と更新された前記入力データとを対応付けて送信する入力データ受渡部とを有し、

前記コンピュータを、

前記複数の処理プログラムの各々を識別する識別情報と、前記複数の処理プログラムの実行順序とを含む実行指示情報を生成する実行指示情報生成部と、

前記ジョブに入力される入力データを入力として最初に実行すべき処理プログラムを実行する引受サーバに、当該入力データと前記実行指示情報とを対応付けて送信する入力データ送信部と

して機能させる制御プログラム。

- [24] 管理サーバにより管理され、利用者の要求するジョブに含まれる複数のステップを処理する複数の処理プログラムのうち、割り当てられた処理プログラムを実行する引受サーバとして、コンピュータを機能させる制御プログラムであって、

前記管理サーバは、

前記複数の処理プログラムの各々を識別する識別情報と、前記複数の処理プログラムの実行順序とを含む実行指示情報を生成する実行指示情報生成部と、

前記ジョブに入力される入力データを入力として最初に実行すべき処理プログラムを実行する引受サーバに、前記入力データと前記実行指示情報とを対応付けて送信する入力データ送信部と

を有し、

前記コンピュータを、

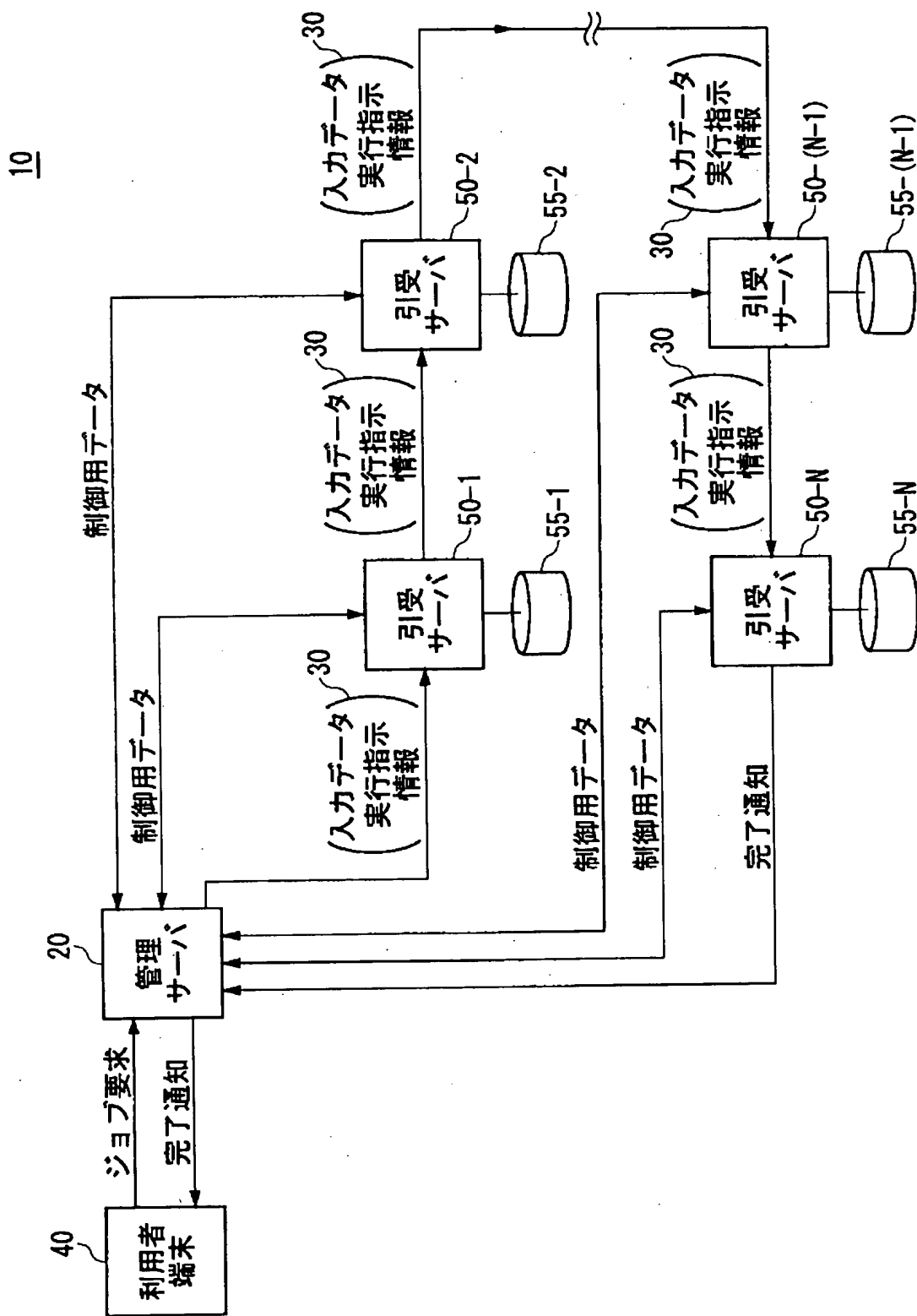
受信した前記入力データを入力として、割り当てられた前記処理プログラムを実行することにより前記入力データを更新する処理プログラム実行部と、

前記入力データに対応付けて受信した実行指示情報に基づいて、更新された前記入力データを入力として次に実行すべき処理プログラムを実行させる他の引受サーバを選択する引受サーバ選択部と、

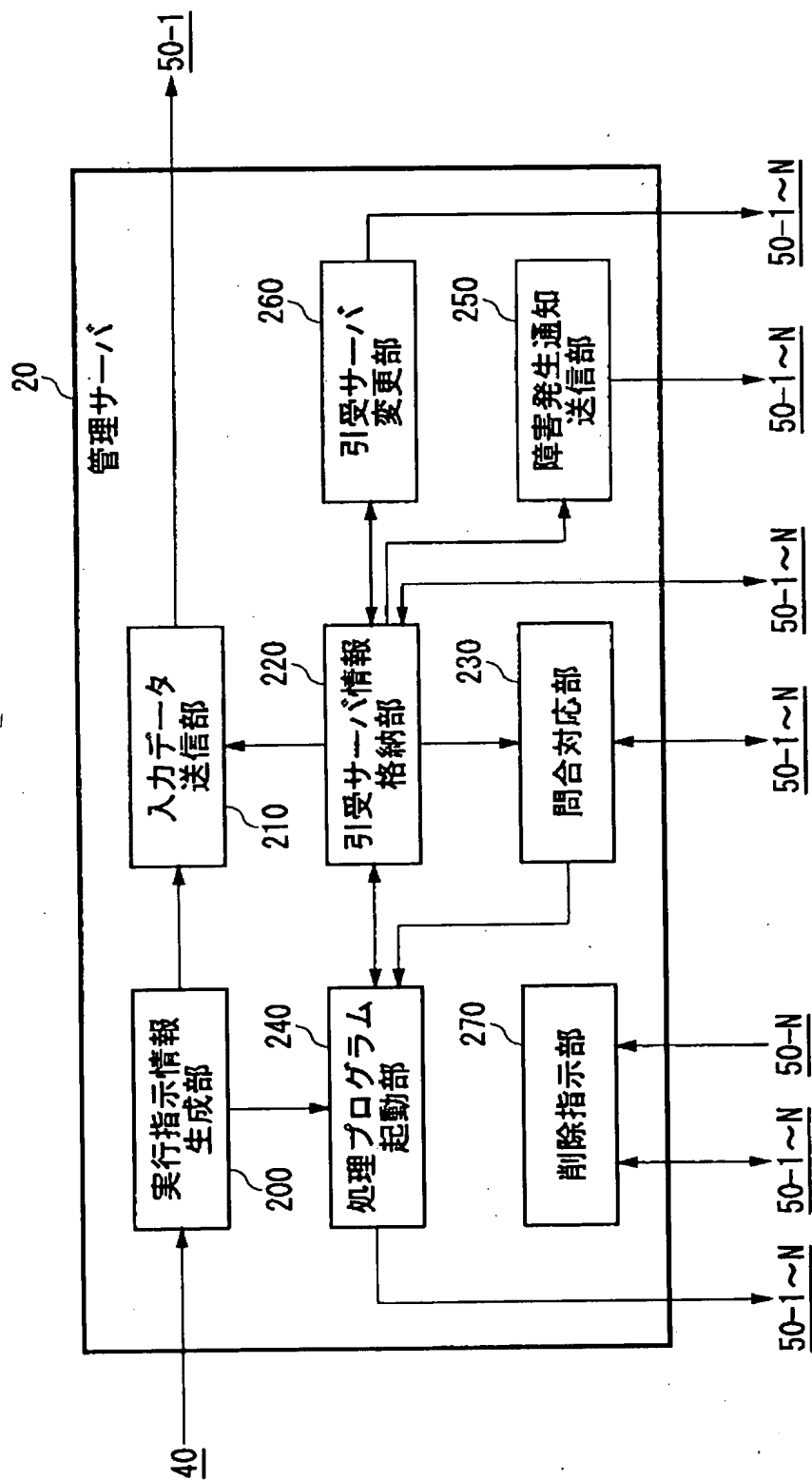
前記引受サーバ選択部により選択された前記引受サーバに、受信した前記実行指示情報と更新された前記入力データとを対応付けて送信する入力データ受渡部として機能させる制御プログラム。

[25] 請求項23又は請求項24記載の制御プログラムを記録した記録媒体。

[図1]

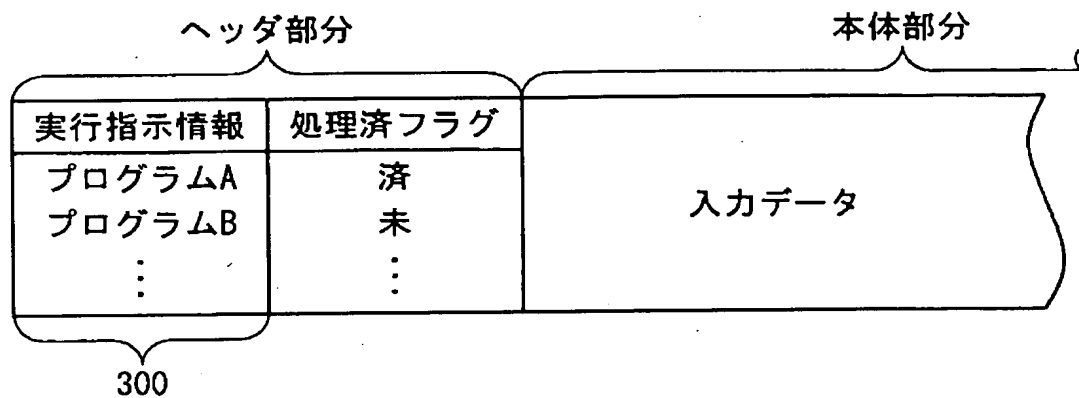


[図2]



[図3]

(a)

30

(b)

300

```
1:  //JOB01      JOB
2:  //STEP01     EXEC  Program=プログラムA
3:  //INdata     FILE  Filename=InfileA
4:  //OUTdata    FILE  Filename=OutfileA
5:  //STEP02     EXEC  Program=プログラムB
6:  //INdata     FILE  Filename=OutfileA
7:  //OUTdata    FILE  Filename=OutfileB
```

⋮  
⋮  
⋮



[図4]

220

引受サーバID	処理プログラム			実行可否情報	最大処理量	使用率
	ID	処理数	最終受信通知			
Node AA	A	1	15:55	可	55	95%
Node BB	B	2	16:15	否	50	60%
Node CC	A	2	14:35	否	55	95%
Node DD	A	2	16:00	否	65	91%
Node EE	—	—	—	—	60	50%
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

[図5]

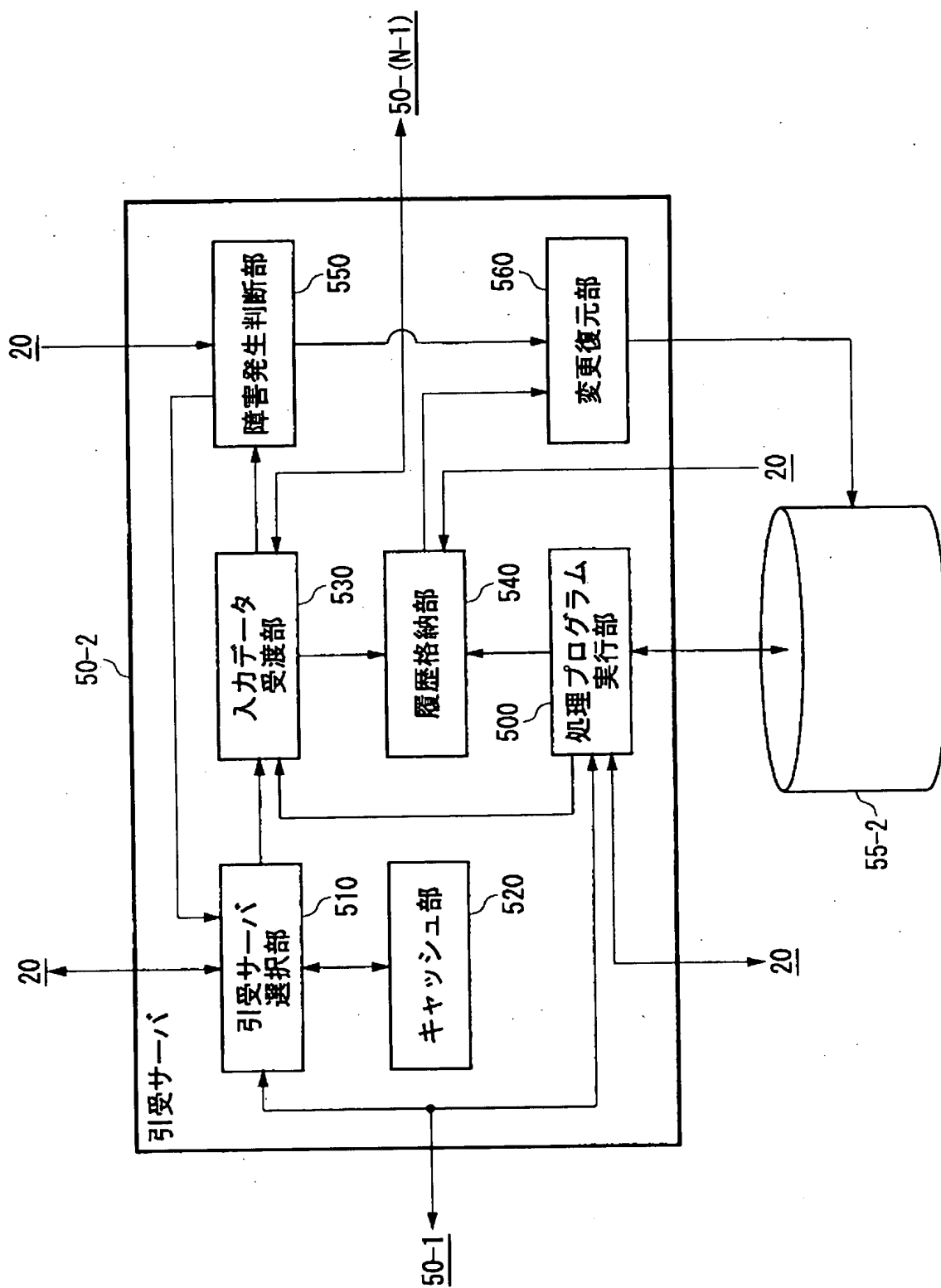
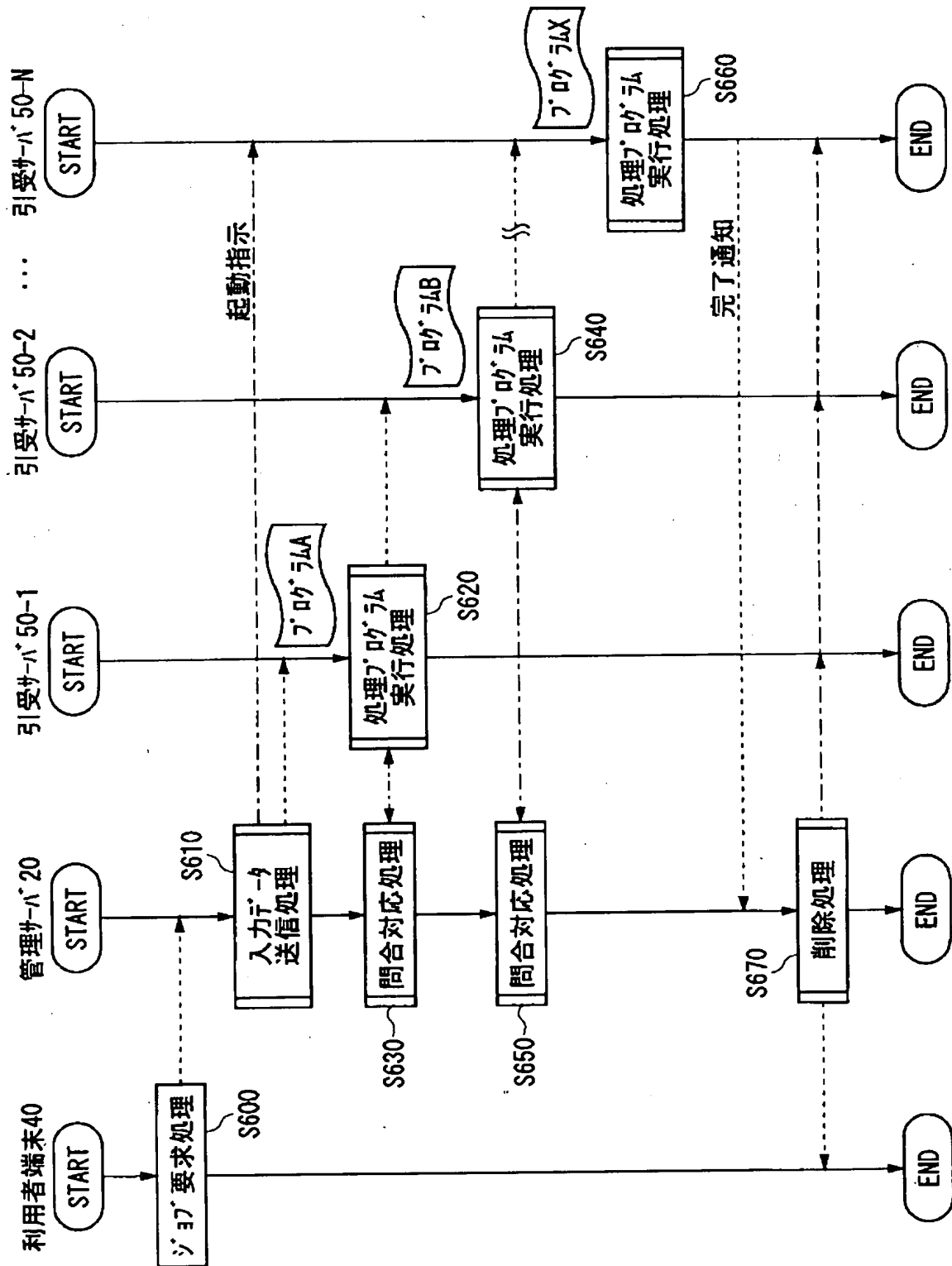
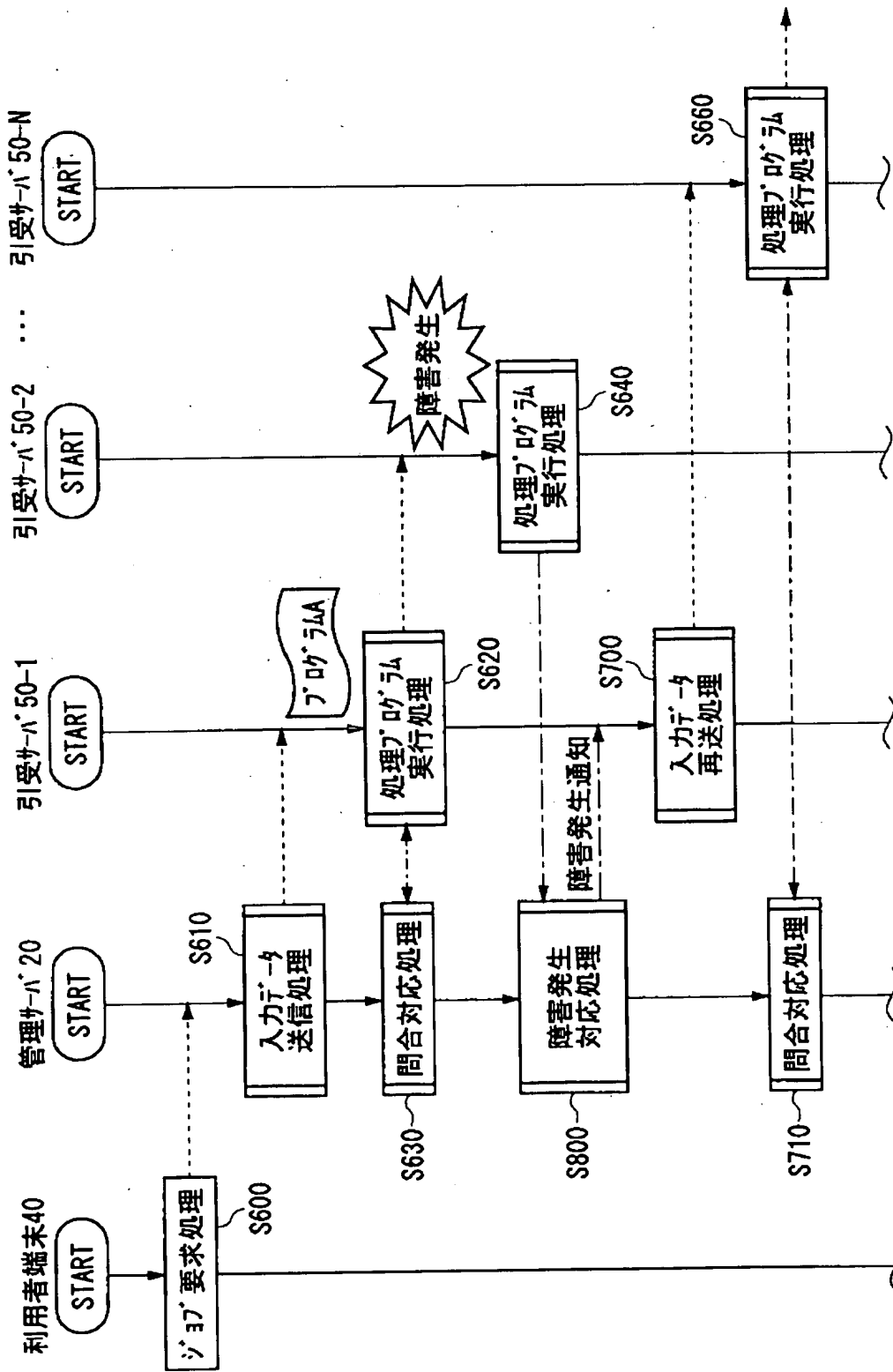


図6

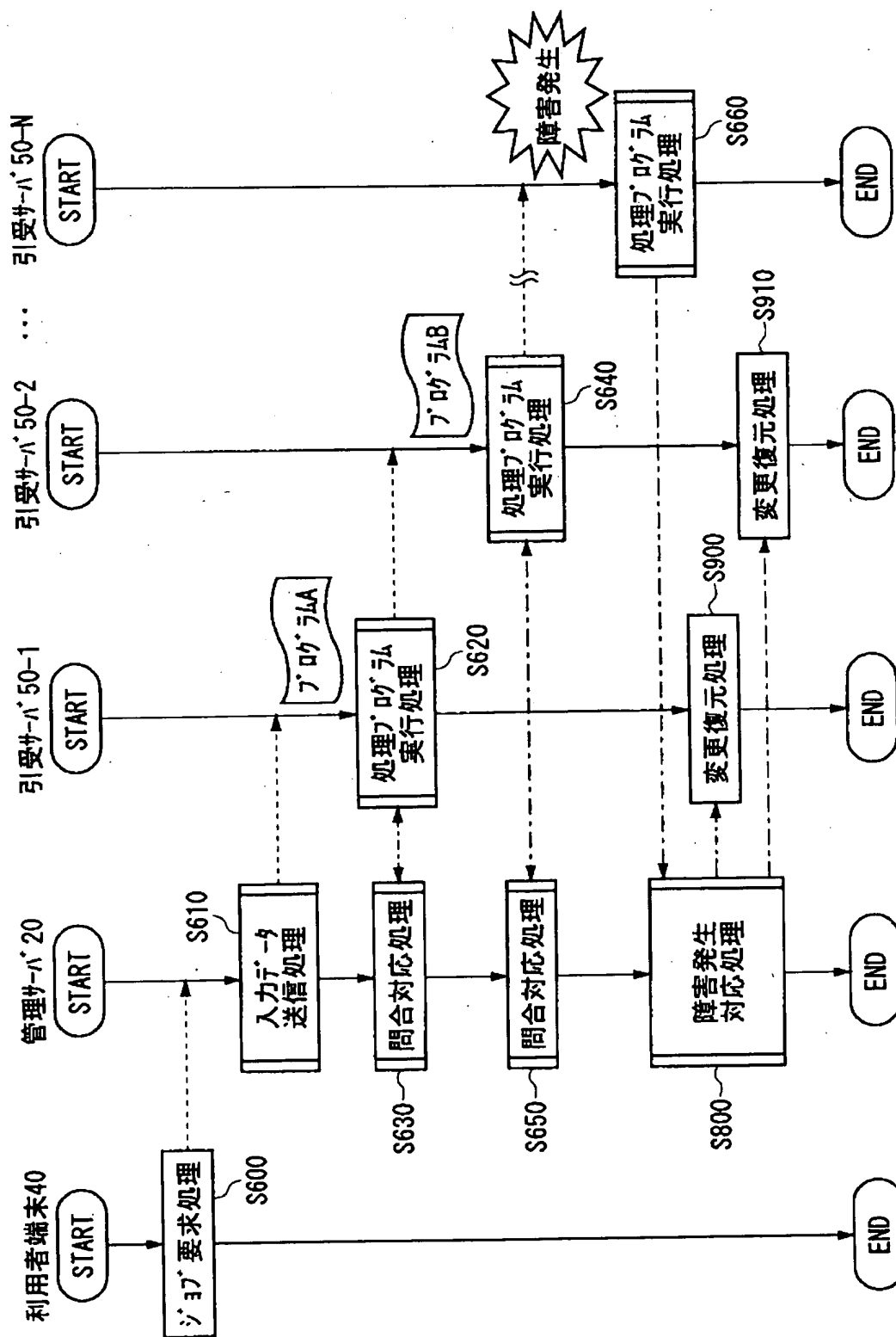




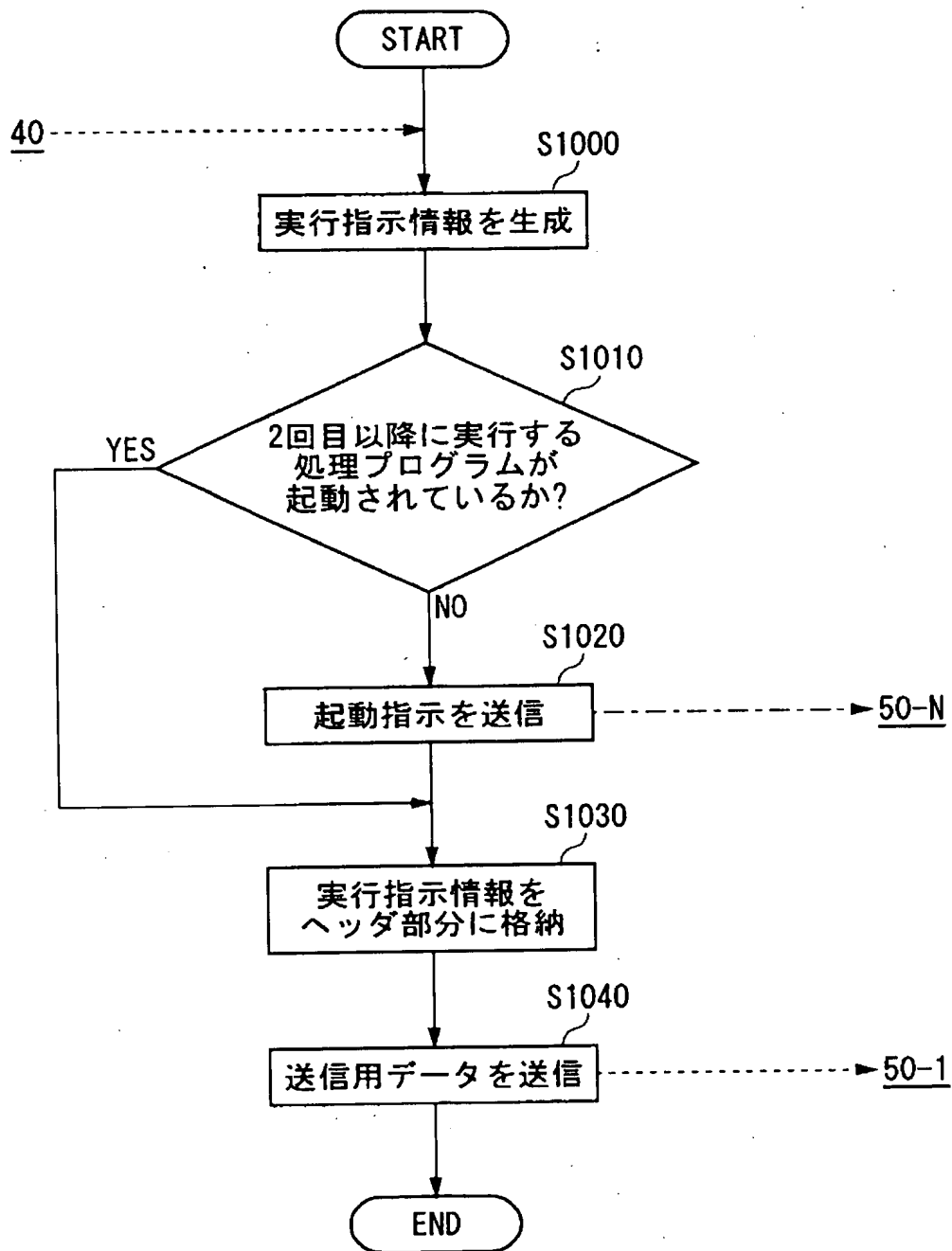
[図8]



[図9]

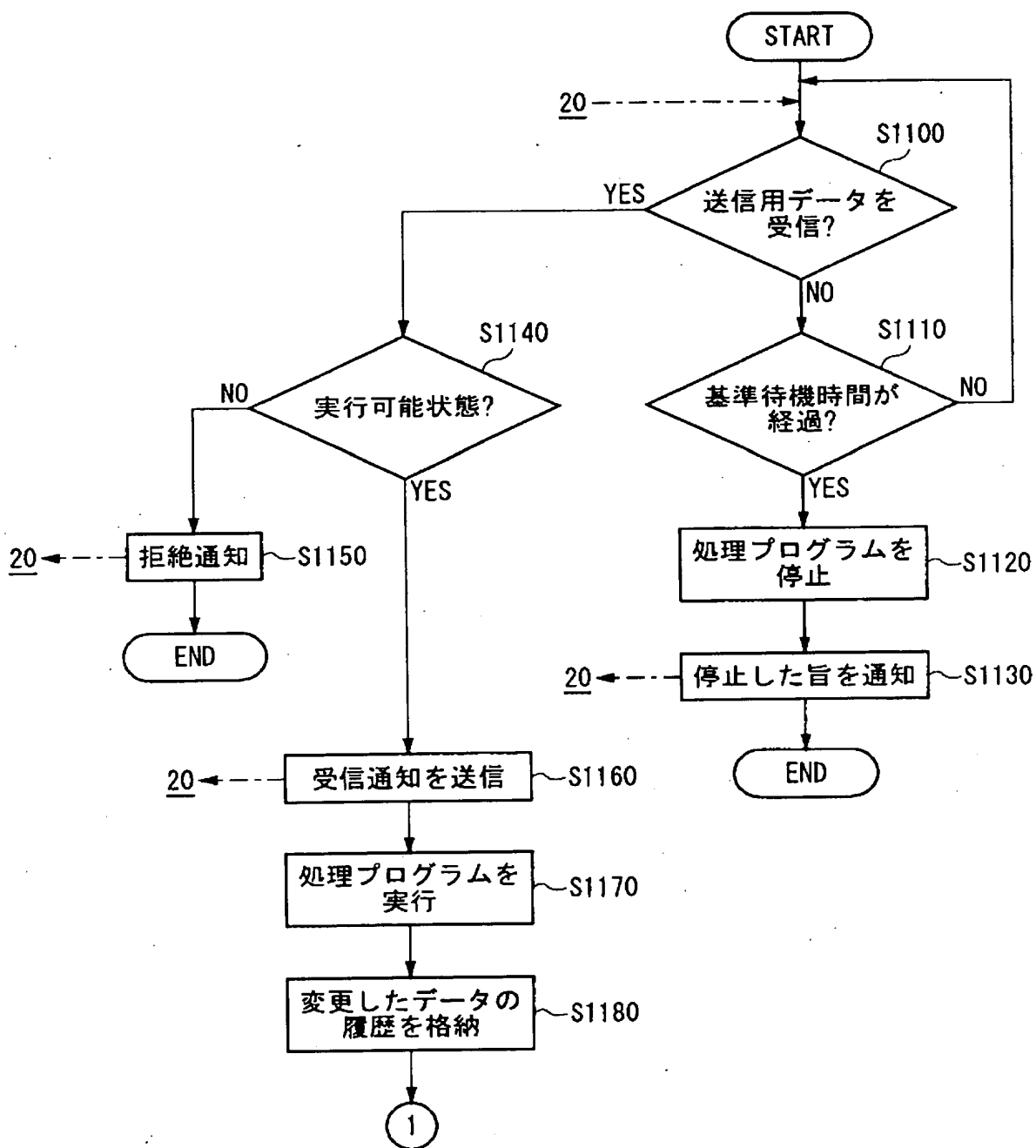


[図10]

S610

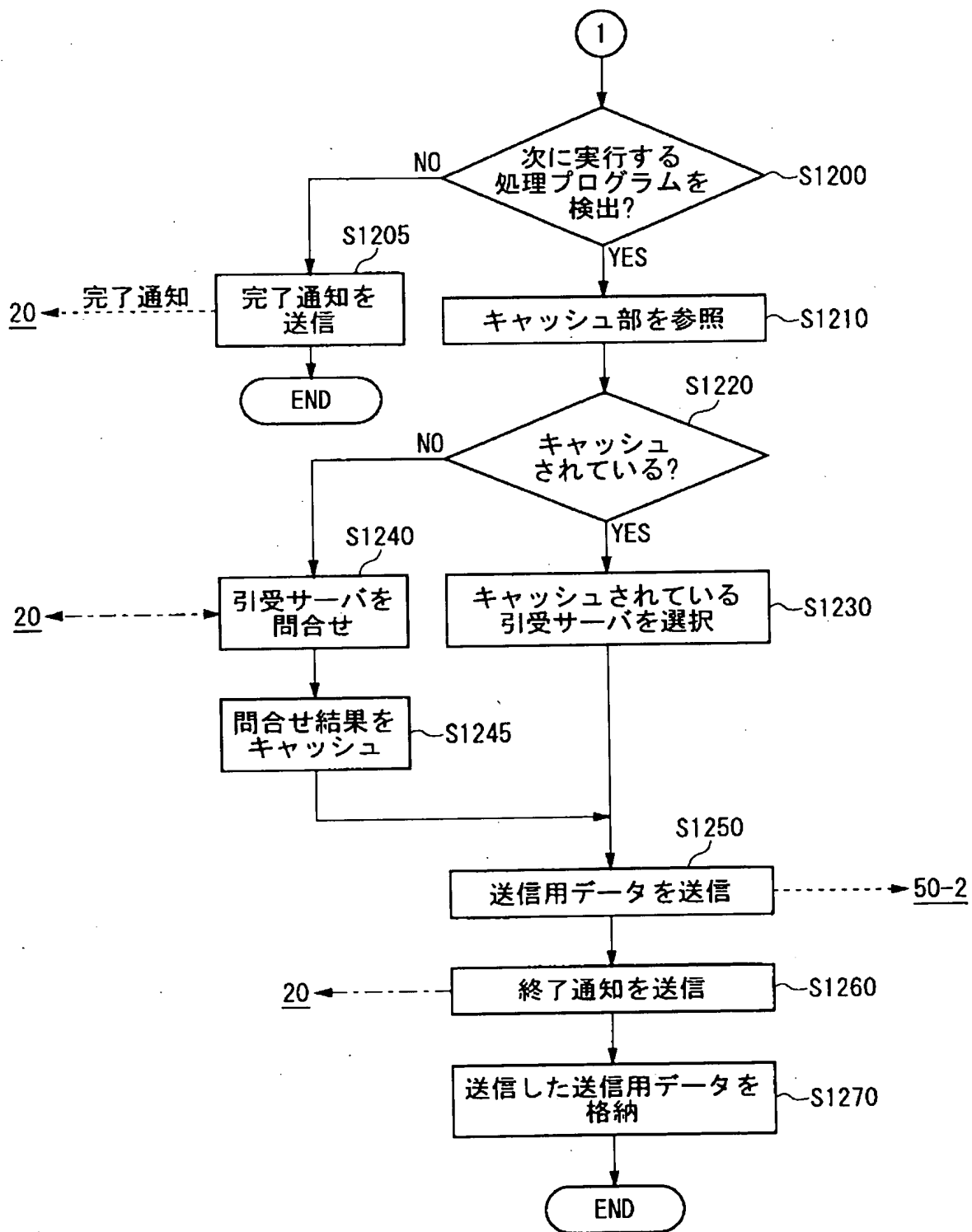
[図11]

S620

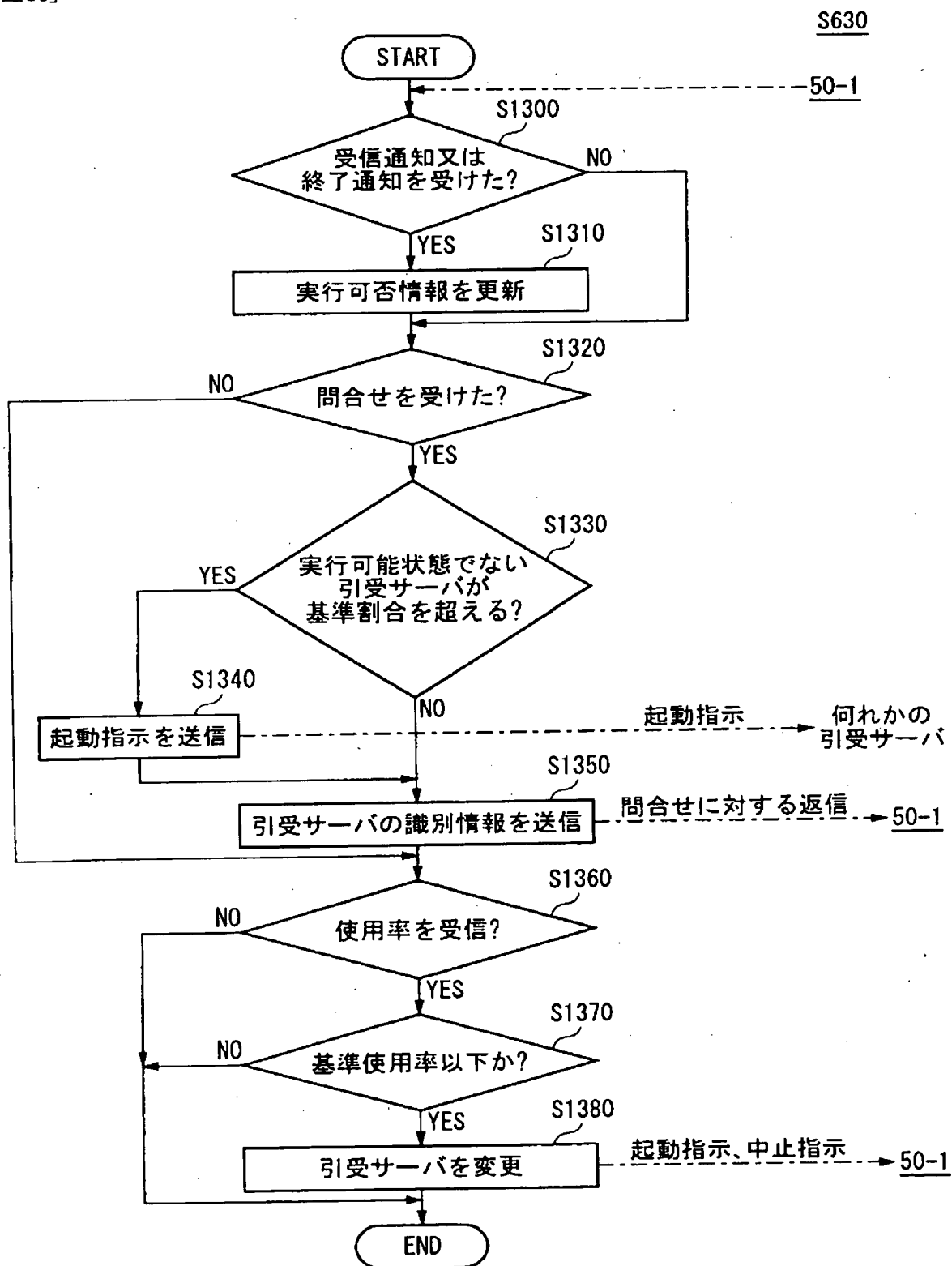




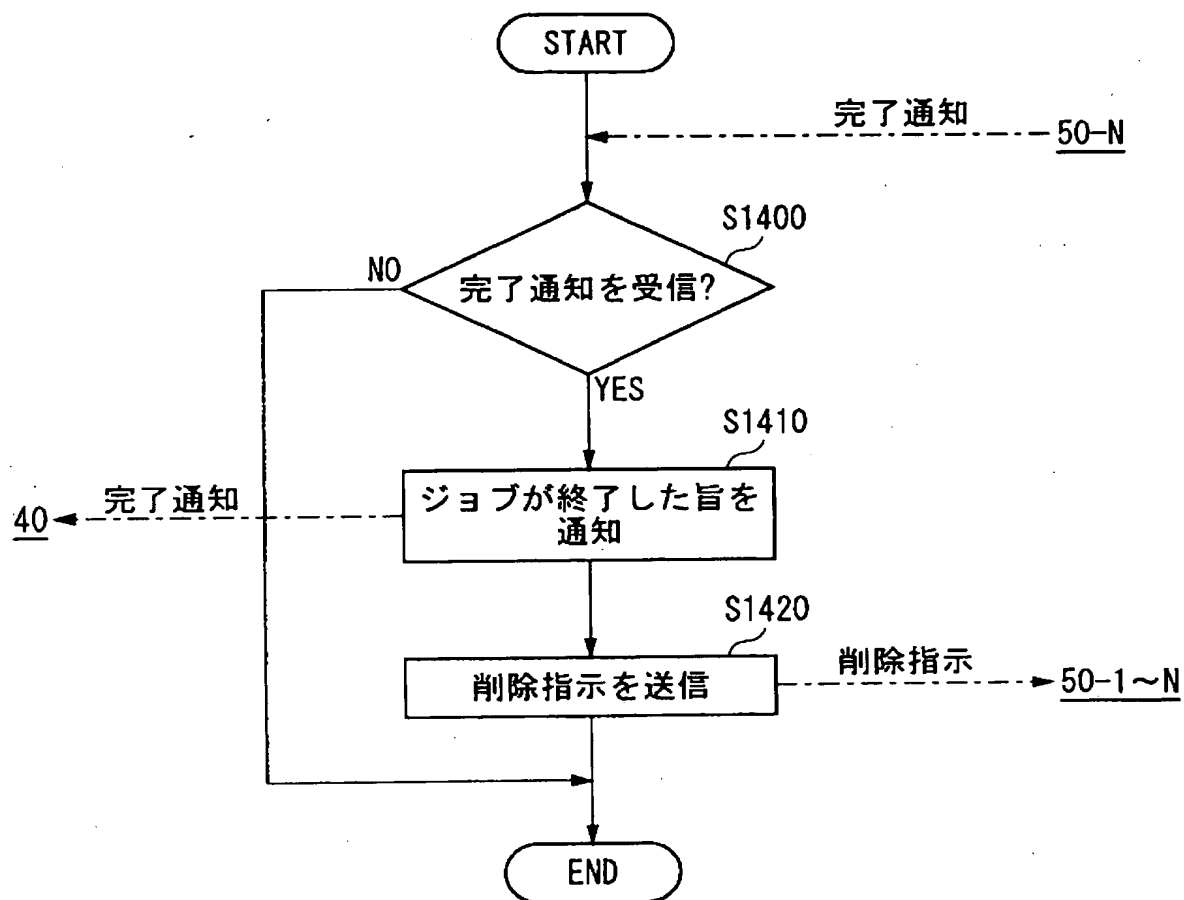
[図12]

S620

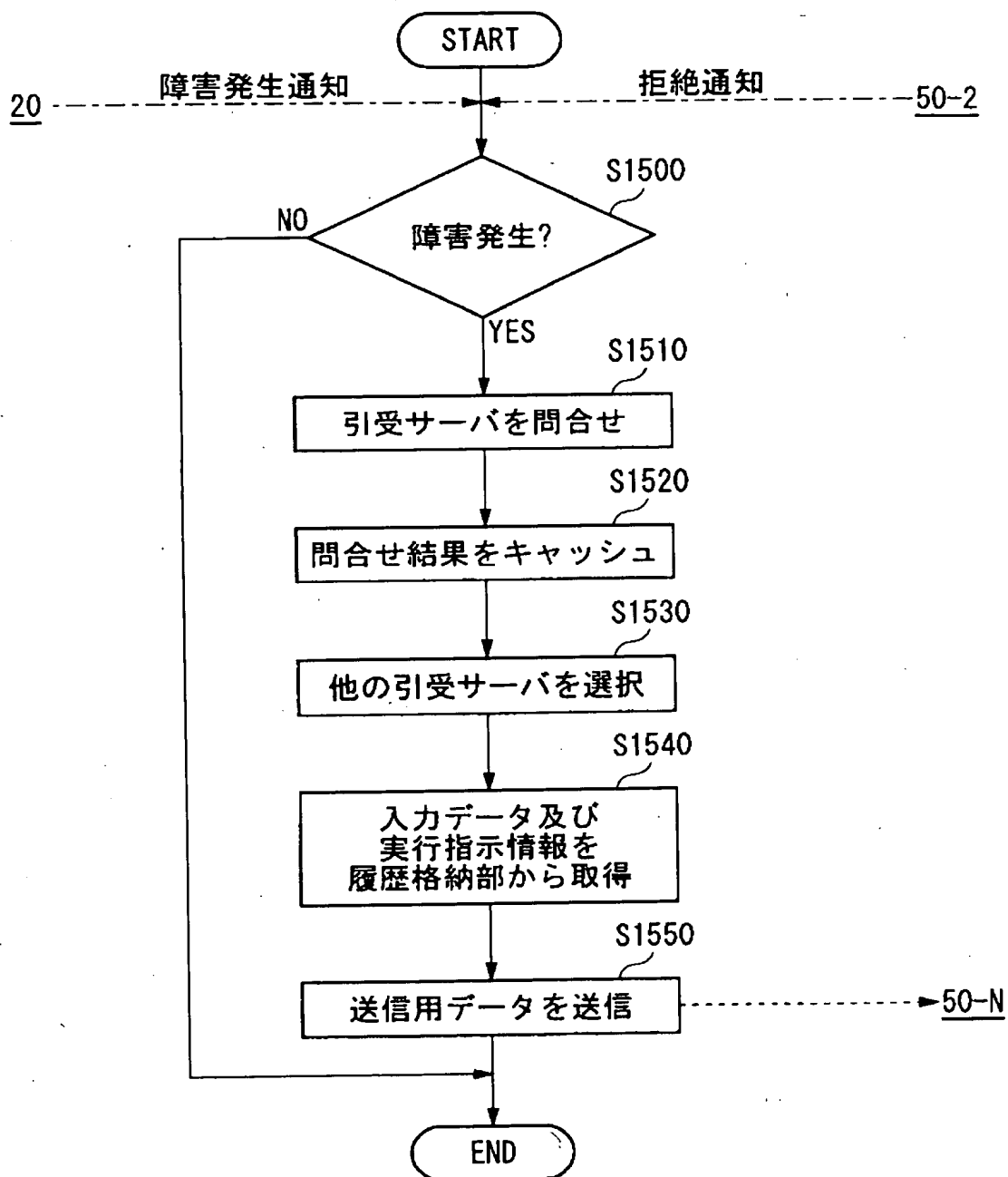
[図13]



[図14]

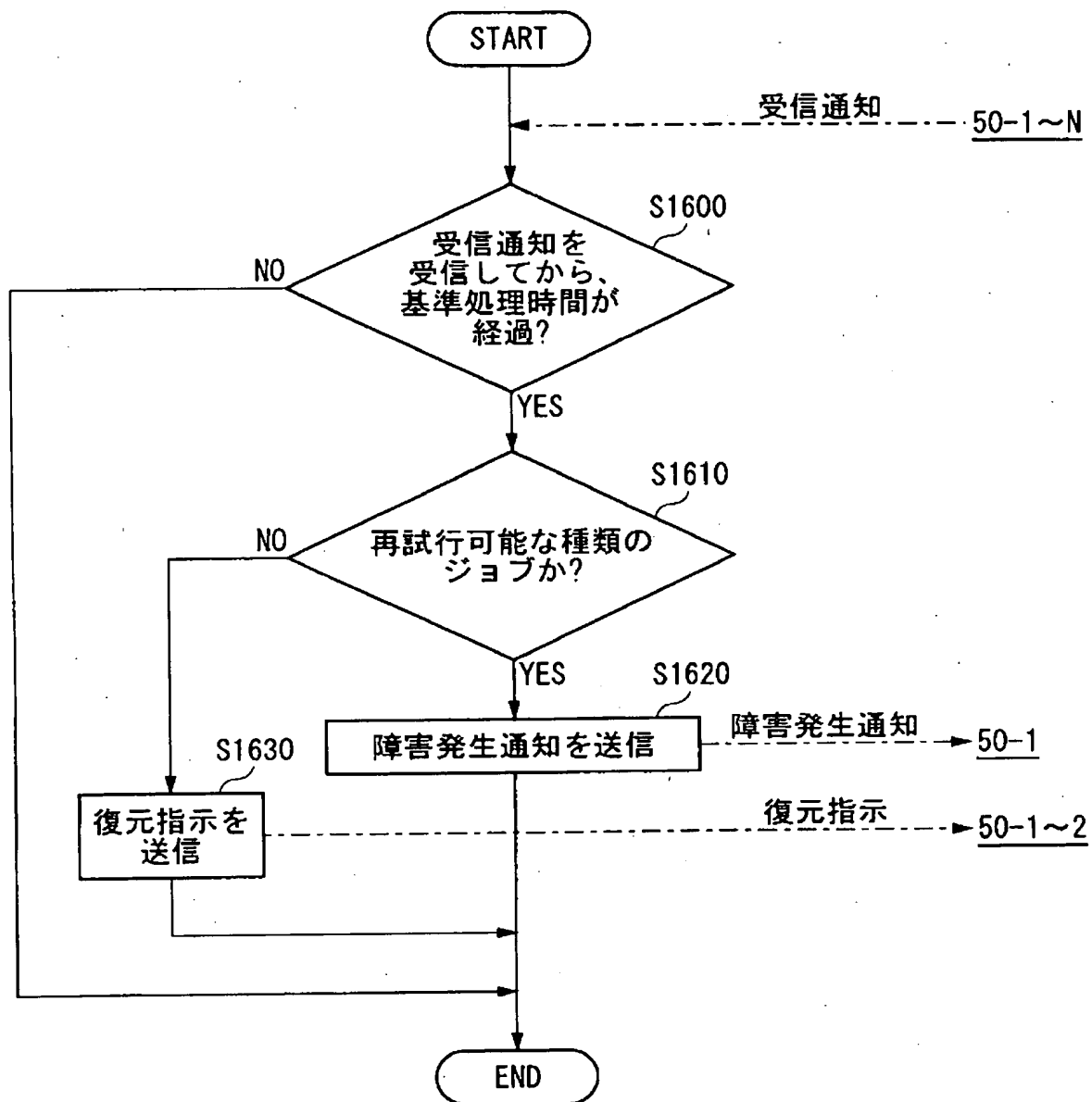
S670

[図15]

S700

[図16]

S800



[図17]

